漢字文献
情報処理研究

第5号

漢字文献情報処理研究会　編
日本中国語CAI研究会　編集協力

好文出版

漢字文獻情報處理研究　第５号

目　次

人文系大学生を対象とした
情報リテラシー教育のための授業プラン

多言語デジタルテキスト、知的生産の方法から
現代社会における基礎知識まで

堀　一成（ほり　かずなり）、山崎　直樹（やまざき　なおき）

I.　はじめに

この文章は、人文科学系の大学生を対象とする情報リテラシー教育の授業プランである。ただし、これは、筆者らの勤務校（外国語大学）で実施することを想定したプランである。そのため、多言語が共存できうる社会において不可欠なリテラシーを重視したものになっていることは否めない。

しかし、できる限り広く人文科学系カリキュラムにおいて採用できる内容にしたつもりである（逆に言えば、「一般の人文科学系」カリキュラムにおいても、この程度の多言語処理能力は必要であるのではなかろうか）。

また、以下の現実的制約に関しては、まだ考慮していない。

A　この授業プランの核になる部分は、全学生に必修の授業として課すことを想定しており、その必修部分は、相当な量の学習項目を含むが、それを、既存のカリキュラムの中に組み込むことが可能かどうか？（既存の教養的教育や専門教育とのバランスはどうなるか？）

B　この授業プランは、核となる部分だけでも、かなり多様な学習項目を含むが、これらを教えられる教師が、相当数（全学生の必修授業をまかなうのに十分なほど）、存在するか？

「考慮していない」意図は、このような現実的な制約にはこだわらず、敢えて、筆者らの勤務校にとって理想と思われる授業を考えてみたかったからである。

しかし、(A)について言えば、我々は、むしろ、このような大規模な基礎教育を提案することにより、大学における教養的教育そのものを作り変えていきたいと考えているので、この大規模な授業プランは、確信犯的な所業だと考えていただきたい。

ただ、「必修」指定にした授業だけを見れば、「90分の授業×通年」という、よくある形態の授業の枠内に納まる。現在、日本の大学において、その教育効果を検証されることなしに広く行われている「第二外国語」と呼ばれる授業科目に、どれだけ時間が割かれているかを考えれば、また、大学教育における授業科目にも、スクラップ＆ビルドが必要とされるなら、筆者らの提案する授業プランも、さほど非現実的なものではあるまい。

II.　授業プランの概要

A.　授業プランの構成

1.　補正教育（単位に算入しない）

2. 基礎的情報リテラシー（必修）
3. 発展的情報リテラシー（選択）
4. 応用リテラシー（選択）

B. 概略

1. 補正教育（第１セメスター，全 15 回）

この授業は、卒業要件単位に算入しない補正教育用の授業である。「補正教育」とは、大学入学時のある分野の学力と、大学教育で必要とされる同分野の学力との差を埋める教育である。

高等学校の普通科でも、「情報」の授業が必修化されるが、各高等学校でのカリキュラムの設定には、その内容の深さと広さの両面で、かなりのばらつきが出ると予想される。

現在、大学の初年次の教育では、英語や数学などにおいて、補正教育の必要が唱えられているが、同様の補正が、情報リテラシー教育においても必要となることを予想し、この授業プランを考えた。

2. 基礎的情報リテラシー：「知的生産のために」（第 2，第 3 セメスター，計 30 回）

全学生に必修として課す授業である。つまり、現代の人文系大学生には欠くことのできない情報リテラシーである、と筆者らが判断した項目を学習する授業であり、今回、提案する授業プランの核になる部分である。

以下のように構成される。

A. Multi-language Digital Text Literacy
B. 情報収集
C. 情報整理
D. 情報利用
E. 情報構築
F. 情報発信

3. 発展的情報リテラシー（第４セメスター，全 15 回）

これは、「必修にするような内容ではないが、専門分野によっては必要となる情報処理技術」という観点で組む授業プランである。

4. 応用リテラシー（第４セメスター，全 15 回）

これは、「情報リテラシー」の授業ではない。現代の人文系大学生にとって必要な「リテラシー」そのものに対する考察を経て組んだ授業プランである。

あるいは、この授業プランを見て、「これは、教養的教育の範疇に進出している」と思われるかもしれないが、それは、むしろ、筆者らの意図どおりである。筆者らは、この授業プランの提案を通して、情報リテラシー教育をも包含した、新しいタイプの包括的な「リテラシー教育」を構築し、従前の大学における「教養的教育」と呼ばれるものを再考してみたいと考えている。

Ⅲ. 補正教育

ここでは、本来高等学校普通科「情報」で既習であるべきレベルに到達していないものに補習の機会を与え、以降の講義の受講者の最低レベルを保証する内容を行う。

【学習項目】

ⅰ. ハードウェアの概要
ⅱ. ネットワークの仕組み
ⅲ. OS とアプリケーション
ⅳ. 代表的なソフトウェアの操作

【目標】

まず、コンピュータシステムのハードウェアの構成（CPU・メモリ・各種ドライブ・ディスプレイ・プリンタなど）について基本的な知識を得る。

ネットワークについても、ネットワークアクセスの概念（プロバイダの必要性など）とインターネットの仕組みを理解する。

ソフトウェアの役割として、全システムの根幹となる OS と、その上で動くアプリケーションソフトウェアの概念を理解する。

Ⅳ 以降で述べる項目を効率よく学習するため

には、最低限使いこなさなくてはならないオフィス系ソフトウェア（OSの基本操作・ワープロ・表計算・Webブラウザ・プレゼンテーションソフト）をスムーズに運用する能力を身に着けている必要がある。到達目標を一般的に記述することは難しいが、現時点では、Microsoft Office Specialist試験の一般レベル合格を目標レベルに設定するのが妥当ではないかと考える。

Ⅳ.　基礎的情報リテラシー:「知的生産のために」

A.　Multi-language Digital Text Literacy

ここでは、以下の「情報収集」「情報整理」「情報発信」などの作業を行う際に必要となるMulti-language Digital Text Literacy（多言語デジタルテキストリテラシー、以下：MDTL）に関する技術と知識を学習する。

【学習項目】

ⅰ. 文字コードとは何か
ⅱ. ローカル文字コード
ⅲ. ユニコード
ⅳ. 表示・入力・「文字化け」

【目標】

まず、コンピュータ上で文字を表示する仕組みを理解する。次に、各地域においてローカルな文字コードが発達したこと、そして、各ローカル文字コードがなぜ共存しにくいのかを理解する。

さらに、ユニコードが必要とされた背景、その思想、ユニコードにおいては、（思想面と実装面の双方において）「文字」がどのように定義されているかを学び、特に、この文字コードが、漢字圏の情報処理に与えた影響について理解する。

そして、これらの理解を元に、システムのデフォールト言語以外の言語の文書を表示する／入力する技術、および、表示／入力のさいの「文字化け」に対処する技術を身につける。

冗談めかして言えば、「多言語環境取扱士」の資格を身につける授業、といったところであろうか。

上述の「技術」は、単なる対症療法的な技術を志向しない。なぜなら、上記のような知識と技術を身につけた者は、その属するコミュニティにおいて（例：学校・職場……）、多言語環境に関するトラブルに対し、説明責任も伴う対処を要求されることになるだろうからである。

B.　情報収集

ここでは、図書館など旧来の情報収集源と、ネットワーク上の情報を横断的に情報源と捕らえ、そこからいかに効率的に情報を得るかについて学ぶ。

【学習項目】

ⅰ. ノートテイキング
ⅱ. ネットワークエチケット
ⅲ. セキュリティ
ⅳ. 検索の技術
ⅴ. ネットワーク情報収集
ⅵ. 図書館・電子ジャーナル

【目標】

まず、すべての学習（情報収集の一形態と考える）の基本として、ノートの取り方・他人の話の聞き方および講師の話の要点のつかみ方・予習復習の仕方などについて実践的に教える。ネットワークを介してスムーズに情報を得るためには一定のマナーを守る必要がある（例えば、掲示板で質問する際にマナーを守らない書き込みをすれば本来得られるはずだった情報を得られないことが有り得る）。メールの文面のマナー・掲示板での発言のマナー・ネットワーク情報の信憑性について考えることなどがあげられよう。

セキュリティに関しては多数の話題があり、どのレベルまで教えるかが難題である。コンピュータウイルスの存在とその防御法、ネットワーク情報の盗聴の可能性と暗号化（メールのデータ・ショッピングサイトでのクレジットカード番号の扱い・無線LANシステムの危険性など）、個人のプ

ライバシーに関する情報を守るにあたって必要な心がけなどが、最低必要な事柄ではないだろうか。

検索の技術の項目は、単なる検索サイト紹介ではない。最小限のコストで自らが必要としている情報を獲得することを目標に、検索内容によって検索サイトとリンク集を使い分けること、検索キーワードの選び方や絞込み検索の方法について学ぶ。絞り込み検索に関連して、論理学の基礎（命題・論理和・論理積・否定など）を簡潔に教えることが必要である。加えて、簡単な正規表現の知識と検索への応用も加えるべきであろう。

ネットワーク上の情報を収集するに際して、上記の検索技術に加え、受講生各自の専門分野でよく使われるWebサイトに関する情報を（専門分野の教員の協力を得て）教える必要がある。

図書館利用リテラシーはネットワーク情報リテラシーと対比させせつつ教育するのが効果的であると思われる。各図書館の基本的な利用法（図書の配架状況、借り方返し方の具体的方法など）に加えて蔵書検索（OPAC）、他大学蔵書検索（Webcat）、国立国会図書館情報、さらに電子ジャーナルの利用法などを教育すべきであろう。

C.　情報整理

ここでは、Bで述べた収集法により集めたデータをどのように整理し、自分の利用目的に合致した形に加工するかを学ぶ。

【学習項目】

ｉ．簡単な統計処理
ii．データマイニング

【目標】

上記の項目では2項目に分割したが、その境界はあいまいである。補正教育において表計算ソフトの簡単な活用法は心得ていることを前提にして、まず、まとまった数のデータの合計・平均・最大値・中央値・最頻値などの基本統計量を得る方法を学習する。

データマイニングとしては、データマイニングの基本概念（多数のデータの中に埋もれた規則性などを発見すること）を紹介した上で、上記の基本統計量に加えて、データの回帰分析・相関係数・主成分分析などの事項を学習させる。併せて、専用のソフトウェアの使用法にも習熟させる。往々にして統計処理をすれば結果が出てくることに安心してしまうことに留意し、数学的厳密さよりも、行おうとしている解析手法が、求める情報に沿ったものであるか検討することの大事さを強調すべきであろう。また、得られた解析結果がなにを意味するかを定性的に捉える訓練も必要である。

D.　情報利用

ここでは、他人の著作物を利用するさいに必要となる知的財産権と、自らの著作を公開するさいに心得ておきたい知的財産権について学ぶ。

【学習項目】

ｉ．著作権法の基礎
ii．学術的研究のための著作権
iii．デジタル著作物の著作権
iv．ネットワーク時代の著作権
ｖ．新しい知的財産権の考え方

【目標】

まず、「著作権法」の基礎について学び、その後、研究目的などでの著作物の利用の注意点、fair useなどの概念について学ぶ。その後、「公衆送信権」「送信化準備権」など、デジタルネットワーク時代に必要な諸概念について理解する。

さらに、近年注目を浴びている、旧来の著作権法の枠組みに囚われない、新しい知的財産権のあり方（例：Creative Commons）などについても学び、他者の著作物を正当に利用することができ、なおかつ、自らの著作物を、その著作物の性格に最もふさわしい権利制限の下に公開することができるようにする。

E.　情報構築（支援ツールの活用も含めて）

ここでは、収集し整理した情報を基に、自ら知的生産物を創出する方法を学ぶ。いわゆる、「論文作法」に類する事柄も含む。しかし、単なる「論

文作法」にとどまらず、PC 上のソフトウェアを積極的に活用して、知的生産の支援を行う術も併せて学ぶ。

【学習項目】

ⅰ. 事実と意見を分ける
ⅱ. 適切な章立て、段落の設定
ⅲ. 適切な引用
ⅳ. 説得をするための論理構築
ⅴ. 参考文献の適切な提示のしかた

【目標】

発想支援ソフト・アウトラインプロセッサ・文献管理ソフトなどの使用法に習熟し、それらを活用して、より高度な生産を効率的に行う技術を身につける。

F. 情報発信

知的生産物を発信するために必要な、知識と技術、およびエチケットを学ぶ。ここで重要なのは、「相手に分かってもらうためのサービス精神」と「相手の環境を考慮する気配り」を身につけることである。

全体は、以下の構成をとる。

1. ファイルの形式、交換と共有
2. プリントアウトする文書の体裁
3. Web サイトの構築
4. プレゼンテーション

1. ファイルの形式、交換と共有

【学習項目】

ⅰ. 知的生産物をどのようなファイルフォーマットで表現すべきか――プレーンテキスト・ワープロ文書・PDF・ハイパーテキストなどの長所と短所
ⅱ. ファイルを交換するのにどのようなメディアを選ぶべきか――FD・HD・CD-ROM・DVD・メール添付・FTP

【目標】

ここでは、知的生産物を盛り込むファイルのフォーマットと、それを伝送するメディアについて、それら各種の長所と短所を理解し、必要に応じて、最もふさわしい形式を選べるようにする。同時に、受け取る相手の環境を考慮することの重要性を学ぶ。

2. プリントアウトする文書の体裁

ここでは、プリントアウトする文書の体裁を効果的に整える技法を学ぶ。

【学習項目】

ⅰ. 一枚物の配付資料・ビラ・報告書の作りかた
ⅱ.「構造とスタイルの分離」という考えかた
ⅲ. 複数ページに渡る文書の構造化、および効率的な作成法

【目標】

効率よく、印象的に、伝えたい内容を伝えるための印刷文書作成技術を身につける。同時に、ワープロソフトなどが豊富に備えている文書作成支援機能を使いこなせるようにする。例えば、「スタイル設定」「索引の作成」「目次の生成」など。

3. Web サイトの構築

ここでは、簡単な Web サイトの構築の実習を行う。

【学習項目】

ⅰ. ハイパーテキストとは何か
ⅱ. 文書を構造化するということ
ⅲ. 構造とスタイルの分離、統一的デザイン
ⅳ. 多言語混在の方法
ⅴ. サイトの階層とナビゲーション
ⅵ. バリアフリー、ユニバーサルデザイン
ⅶ. Web サイトアクセシビリティ⑴：テキストブラウザで閲覧したらどうなるか
ⅷ. Web サイトアクセシビリティ⑵：音声

ブラウザで「閲覧」したらどうなるか

【目標】

「よいWebサイトとは何か」を実習を通して身につける。効率的な情報の提供という観点の重要性とともに、閲覧者がどのような環境下にあっても閲覧できるような配慮が重要であることを理解する。

4.プレゼンテーション

【学習項目】

ｉ．プレゼンテーションの構成のしかた
ⅱ．配付資料について注意すべきこと

【目標】

より効果的なプレゼンテーションのしかたを身につける。

具体的には、Organization（内容をどう構成するか）、Delivery（人前でどのように話すか）、Visual Aids（視覚的効果をどう援用するか）……などの実践的技術について学ぶ。

同時に、PC上のプレゼンテーションツールの使用法に習熟することを目標とする。

Ⅴ．　発展リテラシー

以下では、より高度なリテラシー、より高度な情報処理技術を身につけるための授業プランを、ごく大雑把な形ではあるが、考えてみた。これは、学生が､必要に応じて（自己の専門分野に応じて）選択する性格の授業である。

A.　フィロロジストに必要なリテラシー

【学習項目】

ｉ．Perlなどのスクリプト言語による文字列の操作（検索／抽出／ソート／整形／出力）
ⅱ．形態素分析ツールなどによる解析、KWICコンコーダンスの作成、Nグラムモデルによる解析
ⅲ．計量的文献研究に必要な統計処理技術

B.　構造化文書に関するより高度な知識

【学習項目】

ｉ．マークアップ言語とは何か
ⅱ．HTMLのより高度な理解
ⅲ．XMLの基礎知識

C.　教師に必要なリテラシー

【学習項目】

ｉ．テスティングの理論
ⅱ．成績処理に必要な統計処理

Ⅵ．　応用リテラシー

これは、「情報リテラシー」にとどまらず、現代の人文系大学生に必要な「リテラシー」そのものを身につけさせる授業である。

A.　メタリテラシー

「リテラシー」そのものについて考える授業である。以下のように構成される。

【学習項目】

1. 伝統的リテラシーの欠如
2. デジタルディバイド
3. 多言語化する社会におけるデフォールト言語のリテラシー
4. 多言語対応社会になるための条件

1.伝統的リテラシーの欠如

【学習項目】

ｉ．非識字者への抑圧
ⅱ．現代の問題としての識字問題

【目標】

伝統的なリテラシーの欠如、つまり、読み書き能力の欠如は、どのように、人を抑圧する材料として用いられてきたかを理解する。併せて、現代の日本でさえも「読み書き能力の欠如」は、決し

て過去の問題ではないこと（部落差別の結果としての非識字者の問題は、未だ過去のものではない）を学ぶ。

また、学習障害（例：難読症 dislexia）が原因で、リテラシー能力に欠損があるばあいでも、その障害の実態が社会的にまだ認知されていないゆえに、不利益をこうむっている事例もあることも学ぶ。

2.デジタルディバイド（Digital divide）

デジタルディバイドとは、パソコンやインターネット等の情報技術を利用する能力や、それにアクセスする機会を持つ者と持たない者との間に、情報格差、およびそれに起因するさまざまな格差が生じるとされる問題のことである。

【学習項目】

ⅰ．情報リテラシーを身に付ける機会の欠如
ⅱ．情報リテラシーの欠如がもたらすもの。
ⅲ．デジタルディバイドがもたらすもの。

【目標】

まず、「情報リテラシーの欠如」が、かつての「読み書き能力の欠如」のように、人を抑圧する材料になりかねない現状について学ぶ。

さらに、デジタルディバイドが深刻な問題になりうるのは、情報リテラシーを身につけており、かつ情報技術を活用できる経済環境にある者が、社会の構成員としてのデフォールトな存在であることを前提として、社会体制が構築されつつあることが一因であることなどを理解する。

3.多言語化する社会におけるデフォールト言語のリテラシー

【学習項目】

ⅰ．急速に多言語化する日本社会
ⅱ．多言語対応社会
ⅲ．デフォールト言語のリテラシーの欠如

日本語のリテラシーが不十分な構成員が日本の社会に占める割合は、急速に高くなりつつある。そのような構成員が、日本社会のデフォールト言語（日本語）のリテラシーが不十分であると、かつて非識字者が抑圧されたごとく、そのリテラシーの欠如により、さまざまな抑圧を受けることになる。ここでは、その現状について理解する。

4.多言語対応社会になるための条件

【目標】

複数の言語の話者の存在を許容でき、そして、複数の言語で各種公共サービスを提供できる社会を実現するためには、何が必要か、とくに、そのような社会の実現のためには、我々のリテラシーというものに対する認識をどう変えていかねばならないかを考える。

B.　社会科学系リテラシー

受講生が卒業後に社会人として働いていくに際して、多様な進路のいずれにおいても必要となりそうな基本知識を教える。主な柱として経済・経営学的知識と、法学的知識の獲得を挙げることができる。

1.経済・経営学的知識

特に責任ある立場の社会人として活躍するためには、あらゆる分野において経営的視点が重要になる。以下に列挙するような経営学の基本項目は進む分野にかかわらず教養として必要な事項であると考える。

【学習項目】

ⅰ．会計の基礎
ⅱ．経営戦略
ⅲ．業務管理

【目標】

まず、会計の基礎として複式簿記の基本原理の理解と財務諸表の意味について学ぶ。

続いて経営戦略の項目として、その財務諸表に書かれた数値から経営状態などの把握をするための諸表分析について学ぶ。最近は会計ソフトの機能が充実し、細かな計算を正確に行う能力より、その結果から有意な情報を読み取る能力がより重

要と思われる。さらに読み取った情報をとるべき行動の決定の材料にするため、意思決定理論の基本を学ぶべきであろう。前記のデータマイニングの知識も動員すべきである。

仕事を始め物事は順調に進まない方が普通と思っておいたほうが良い。その被害を最小限にするために、業務管理の様々なテクニック（リスクマネージメント・業務行程管理）も学ぶ必要がある。併せて、専用のソフトウェアの使用法にも習熟させる。

2. 法学的知識

1.に挙げた会計的知識と同様、社会人として活躍するためには、基本的な法学的知識が不可欠である。それは、自らの発展の礎となるものでもあり、自らを守るものでもある。以下に列挙するような法律の基本項目は教養として必要な事項であると考える。

【学習項目】

i．民法、商法
ii．知的財産権、特許法
iii．労働関係の法律、社会保険・年金関係の法律
iv．刑法、不正アクセス防止法、PL 法、不正競争防止法、独占禁止法

【目標】

まず、あらゆる産業分野に共通に必要と考えられる民法・商法の基本を学ぶ。

自らの発展のためには知的財産権の確保が重要である。関連の法的知識を学ぶ必要がある。

さらに、自らの収入や労働環境を守るために労働関係の法律や、保険・年金関係の知識が不可欠であろう。

加えて、本来自らの利益になるはずの仕事における活動が、法的知識の不足によってかえって自らの不利益になることがないよう心がけなくてはならない。刑法、不正アクセス防止法、PL 法など職務上の行為が不正なものかどうかを判断する十分な知識を獲得しておくべきである。

Ⅶ. おわりに

以上のように、現時点で人文系学生が身についた知識として持っているべき項目を整理・列挙してみた。この案はまだ現実に講義を行った経験に基づくものではなく、多々不備もあろう。不備な点に気づかれた読者はぜひ著者に意見を送っていただきたい。そしてより良いリテラシー教育のあり方を考える輪が広がればと思う。多くの大学でこのようなプランに基づく統一的なリテラシー教育を行わざるを得ない状況がすぐそこに迫ってきているのではないだろうか。本稿がそのような将来教育の検討の一助となれば幸いである。

Ⅷ. 付録：コースデザイン

最後に、付録として、この授業プランを実施するためのコースデザイン案を以下に掲げる。

A. 必修／選択の別

- 補正教育　（受講は任意・単位化しない）
- 基礎リテラシーA　（全学生必修）
- 基礎リテラシーB　（全学生必修）
- 発展リテラシー　（自由選択）
- 応用リテラシーA　（自由選択）
- 応用リテラシーB　（自由選択）

B. 各セメスターへの配当

セメスター	授業名
1	補正教育
2	基礎リテラシーA
3	基礎リテラシーB
4	発展リテラシー 応用リテラシーA 応用リテラシーB

C.　概要

1. 補正教育（第１セメスター、15 回）

- ●イントロ　⑴
- ●ハードウェアの概要　⑴
- ●ネットワークの仕組み　⑴
- ● OS とアプリケーション　⑴
- ●代表的なソフトウェアの操作　⑽
- ●まとめ　⑴

2. 基礎リテラシーＡ：知的探索（第２セメスター，15 回）

- ● Multi-language Digital Text Literacy　⑵
  文字コードとは何か／ローカル文字コード／ユニコード／表示・入力・「文字化け」
- ●情報収集　⑶
  ネットワークエチケット／セキュリティ／検索の技術／ネットワーク情報収集／図書館／電子ジャーナル
- ●情報整理　⑺
  簡単な統計処理／データマイニング
- ●情報利用　⑵
  著作権法の基礎／学術的研究のための著作権／デジタル著作物の著作権／ネットワーク時代の著作権／新しい知的財産権の考え方
- ●まとめ　⑴

3. 基礎リテラシーＢ：知的生産（第３セメスター、15 回）

- ●情報構築　⑷
  事実と意見を分ける／適切な章立て、段落の設定／適切な引用／説得をするための論理構築／参考文献の適切な提示のしかた
- ●情報発信　⑽
  ファイルの形式、交換と共有／プリントアウトする文書の体裁／ Web サイトの構築／プレゼンテーション
- ●まとめ　⑴

4. 発展リテラシー（第４セメスター、15 回）

- ●フィロロジストに必要なリテラシー　⑻
  Perl などのスクリプト言語による文字列の操作／形態素分析ツールなどによる解析、KWIC コンコーダンスの作成、N グラムモデルによる解析／計量的文献研究に必要な統計処理技術
- ●構造化文書に関するより高度な知識　⑶
  マークアップ言語とは何か／ HTML のより高度な理解／ XML の基礎知識
- ●教師に必要なリテラシー　⑶
  テスティングの理論／成績処理に必要な統計処理
- ●まとめ　⑴

5. 応用リテラシーＡ：メタリテラシー（第４セメスター、15 回）

伝統的リテラシーの欠如／デジタルディバイド／多言語化する社会におけるデフォールト言語のリテラシー／多言語対応社会になるための条件

6. 応用リテラシーＢ：社会科学系リテラシー（第４セメスター、15 回）

経済・経営学的知識／法学的知識

中国古籍における文字出現頻度および分布に関する統計と分析

朱　岩（しゅ　がん）
訳：千田　大介（ちだ　だいすけ）

はじめに

2003年、中国教育部言語情報局による「中国古籍における文字出現頻度と古籍漢字の符号化文字集合における分布統計」プロジェクトの要請により、文字数・収録作品ともに最も膨大な代表的叢書『四庫全書』・『四部叢刊』における漢字の使用状況について統計作業を行った。ここに、関連する統計結果と、それに対する初歩的な分析を紹介し、ひろく学術界の参考に供するものである。

統計に用いたコーパス

- 『文淵閣四庫全書』
  収録作品：3,500余部、36,000余冊
  総文字数：698,076,596文字（書中の僻字を除く）
- 『四部叢刊』
  収録作品：504部、3,134冊
  総文字数：90,616,538文字

以上の二部の叢書はいずれも完全にテキストデータ化されている。

統計に用いた符号化文字集合

- ISO/IEC 10646-1:2000 / Unicode3.1、CJK+
  『四庫』および『四部』のデジタル版で採用された符号化文字集合でもある。

統計方法

自作の統計ソフトをもちいた。

1.　漢字出現頻度一覧（降順）

A. 上位120字一覧（120字以下は省略）

次ページ参照

B. 統計中の文字に関するパラメータ

解説

- No:00001：
  文字出現頻度の順位
- U+04E4B：
  UnicodeのBMPにおけるコードポイント
- 0019891628：
  「之」のコーパス中における出現回数
- 252.2099‱：
  「之」のコーパス中における出現頻度（万分率—単位：‱）
- 2.5221%
  「之」のコーパス中における出現頻度（パーセント）

No:00001　U+04E4B

0019891628
252.2099‱
2.5221%

No:00001 U+04E4B 之 0019891628 252.2099‰ 2.5221%	No:00002 U+04E0D 不 0009733996 123.4193‰ 3.7562%	No:00003 U+04EE5 以 0009480560 120.2059‰ 4.9583%	No:00004 U+04E5F 也 0007881524 99.9314‰ 5.9576%	No:00005 U+0800C 而 0007697614 97.5996‰ 6.9336%	No:00006 U+05176 其 0006896797 87.4458‰ 7.8081%	No:00007 U+04EBA 人 0006750929 85.5964‰ 8.6640%	No:00008 U+070BA 為 0006615021 83.8731‰ 9.5028%	No:00009 U+06709 有 0006209740 78.7345‰ 10.2901%	No:00010 U+08005 者 0005587903 70.8501‰ 10.9986%
No:00011 U+04E00 一 0005279990 66.9460‰ 11.6681%	No:00012 U+066F0 曰 0005235232 66.3785‰ 12.3319%	No:00013 U+05B50 子 0004579575 58.0653‰ 12.9125%	No:00014 U+065BC 於 0004305732 54.5932‰ 13.4585%	No:00015 U+05341 十 0004185422 53.0678‰ 13.9891%	No:00016 U+05927 大 0003590583 45.5257‰ 14.4444%	No:00017 U+06240 所 0003516804 44.5902‰ 14.8903%	No:00018 U+04E8C 二 0003437197 43.5809‰ 15.3261%	No:00019 U+04E09 三 0003338241 42.3262‰ 15.7494%	No:00020 U+04E2D 中 0003184692 40.3793‰ 16.1532%
No:00021 U+07121 無 0003175745 40.2659‰ 16.5558%	No:00022 U+04E28 丨 0003166824 40.1528‰ 16.9573%	No:00023 U+05E74 年 0003163339 40.1086‰ 17.3584%	No:00024 U+05247 則 0003053478 38.7156‰ 17.7456%	No:00025 U+04E0B 下 0002960595 37.5379‰ 18.1210%	No:00026 U+05929 天 0002919091 37.0117‰ 18.4911%	No:00027 U+04E0A 上 0002887089 36.6059‰ 18.8571%	No:00028 U+06B64 此 0002834051 35.9335‰ 19.2165%	No:00029 U+0738B 王 0002754461 34.9243‰ 19.5657%	No:00030 U+066F8 書 0002619030 33.2072‰ 19.8978%
No:00031 U+0662F 是 0002602664 32.9997‰ 20.2278%	No:00032 U+0516C 公 0002576421 32.6669‰ 20.5545%	No:00033 U+04E8B 事 0002519440 31.9444‰ 20.8739%	No:00034 U+081EA 自 0002480552 31.4514‰ 21.1884%	No:00035 U+05728 在 0002428095 30.7863‰ 21.4963%	No:00036 U+053EF 可 0002422115 30.7104‰ 21.8034%	No:00037 U+08A00 言 0002368329 30.0285‰ 22.1037%	No:00038 U+081F3 至 0002344512 29.7265‰ 22.4009%	No:00039 U+04E94 五 0002342263 29.6980‰ 22.6979%	No:00040 U+05982 如 0002338972 29.6563‰ 22.9945%
No:00041 U+065E5 日 0002325150 29.4810‰ 23.2893%	No:00042 U+05C71 山 0002299699 29.1583‰ 23.5809%	No:00043 U+056DB 四 0002293069 29.0742‰ 23.8716%	No:00044 U+06545 故 0002268200 28.7589‰ 24.1592%	No:00045 U+06587 文 0002232019 28.3002‰ 24.4422%	No:00046 U+08207 與 0002204117 27.9464‰ 24.7217%	No:00047 U+053C8 又 0002194105 27.8195‰ 24.9999%	No:00048 U+05F97 得 0002190762 27.7771‰ 25.2776%	No:00049 U+06642 時 0002172369 27.5439‰ 25.5531%	No:00050 U+0884C 行 0002106107 26.7037‰ 25.8201%
No:00051 U+05DDE 州 0002027504 25.7071‰ 26.0772%	No:00052 U+077E5 知 0002018460 25.5924‰ 26.3331%	No:00053 U+06708 月 0001995017 25.2952‰ 26.5861%	No:00054 U+05F8C 後 0001966734 24.9366‰ 26.8354%	No:00055 U+05357 南 0001932956 24.5083‰ 7.0805%	No:00056 U+0751F 生 0001893680 24.0103‰ 27.3206%	No:00057 U+0898B 見 0001832363 23.2329‰ 27.5529%	No:00058 U+09053 道 0001789762 22.6927‰ 27.7799%	No:00059 U+04EA6 亦 0001748031 22.1636‰ 28.0015%	No:00060 U+08B02 謂 0001747969 22.1628‰ 28.2231%
No:00061 U+06771 東 0001690479 21.4339‰ 28.4375%	No:00062 U+0541B 君 0001689333 21.4194‰ 28.6517%	No:00063 U+0660E 明 0001688211 21.4051‰ 28.8657%	No:00064 U+076F8 相 0001684876 21.3628‰ 29.0794%	No:00065 U+07136 然 0001674640 21.2331‰ 29.2917%	No:00066 U+06C34 水 0001671171 21.1891‰ 29.5036%	No:00067 U+04ECA 今 0001668277 21.1524‰ 29.7151%	No:00068 U+07686 皆 0001649437 20.9135‰ 29.9242%	No:00069 U+04F5C 作 0001635411 20.7357‰ 30.1316%	No:00070 U+0570B 國 0001631438 20.6853‰ 30.3384%
No:00071 U+0540D 名 0001631149 20.6816‰ 30.5453%	No:00072 U+04F7F 使 0001606696 20.3716‰ 30.7490%	No:00073 U+081E3 臣 0001604215 20.3401‰ 30.9524%	No:00074 U+051FA 出 0001603163 20.3268‰ 31.1556%	No:00075 U+0767E 百 0001600496 20.2930‰ 31.3586%	No:00076 U+0672A 未 0001595486 20.2294‰ 31.5609%	No:00077 U+04E8E 于 0001582654 20.0667‰ 31.7615%	No:00078 U+04F55 何 0001571852 19.9298‰ 31.9608%	No:00079 U+08A69 詩 0001561947 19.8042‰ 32.1589%	No:00080 U+0592A 太 0001550011 19.6529‰ 32.3554%
No:00081 U+0897F 西 0001544859 19.5875‰ 32.5513%	No:00082 U+05B98 官 0001544106 19.5780‰ 32.7471%	No:00083 U+077E3 矣 0001540602 19.5336‰ 32.9424%	No:00084 U+0516D 六 0001529457 19.3923‰ 33.1363%	No:00085 U+07528 用 0001524024 19.3234‰ 33.3296%	No:00086 U+06B63 正 0001495257 18.9586‰ 33.5191%	No:00087 U+080FD 能 0001486162 18.8433‰ 33.7076%	No:00088 U+07E23 縣 0001455744 18.4576‰ 33.8922%	No:00089 U+0592B 夫 0001445475 18.3274‰ 34.0754%	No:00090 U+04E91 云 0001406222 17.8297‰ 34.2537%
No:00091 U+06C0F 氏 0001385551 17.5676‰ 34.4294%	No:00092 U+05730 地 0001375207 17.4365‰ 34.6038%	No:00093 U+0540C 同 0001372766 17.4055‰ 34.7778%	No:00094 U+05143 元 0001363996 17.2943‰ 34.9508%	No:00095 U+058EB 士 0001358876 17.2294‰ 35.1231%	No:00096 U+06210 成 0001343501 17.0345‰ 35.2934%	No:00097 U+0967D 陽 0001340765 16.9998‰ 35.4634%	No:00098 U+0975E 非 0001335231 16.9296‰ 35.6327%	No:00099 U+065B9 方 0001325011 16.8000‰ 35.8007%	No:00100 U+05FC3 心 0001319258 16.7271‰ 35.9680%
No:00101 U+05165 入 0001302261 16.5116‰ 36.1331%	No:00102 U+09577 長 0001296295 16.4359‰ 36.2975%	No:00103 U+053CA 及 0001287909 16.3296‰ 36.4608%	No:00104 U+05B89 安 0001281204 16.2446‰ 36.6232%	No:00105 U+050B3 傳 0001265496 16.0454‰ 36.7837%	No:00106 U+07576 當 0001251855 15.8725‰ 36.9424%	No:00107 U+08AF8 諸 0001243065 15.7610‰ 37.1000%	No:00108 U+05148 先 0001240383 15.7270‰ 37.2573%	No:00109 U+05DF2 已 0001216035 15.4183‰ 7.4114%	No:00110 U+079AE 禮 0001214983 15.4050‰ 37.5655%
No:00111 U+098A8 風 0001210352 15.3463‰ 37.7190%	No:00112 U+04E5D 九 0001206422 15.2964‰ 37.8719%	No:00113 U+05E1D 帝 0001206189 15.2935‰ 38.0249%	No:00114 U+04E43 乃 0001193392 15.1312‰ 38.1762%	No:00115 U+05C07 將 0001187450 15.0559‰ 38.3267%	No:00116 U+08ECD 軍 0001180587 14.9689‰ 38.4764%	No:00117 U+05317 北 0001175293 14.9017‰ 38.6254%	No:00118 U+057CE 城 0001165871 14.7823‰ 38.7733%	No:00119 U+082E5 若 0001164704 14.7675‰ 38.9209%	No:00120 U+05E73 平 0001159674 14.7037‰ 39.0680%

2. 文字出現頻度表の漢字カバー率統計

文字種数	カバー率
0	0.00%
10	11.00%
20	16.15%
30	19.90%
40	22.99%
50	25.82%
60	28.22%
70	30.34%
80	32.36%
90	34.25%
100	35.97%

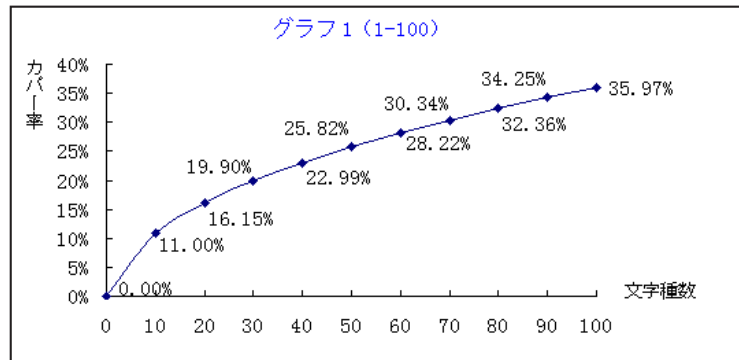
表 1

文字種数	カバー率
100	35.97%
150	43.15%
200	48.71%
250	53.15%
300	56.83%
350	59.99%
400	62.72%
450	65.08%
500	67.19%

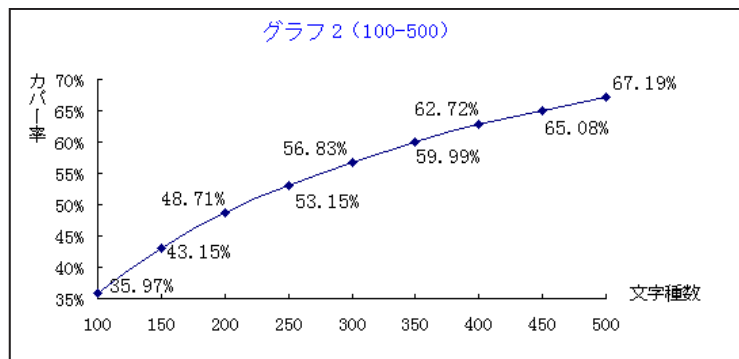
表 2

文字種数	カバー率
500	67.19%
600	70.88%
700	74.00%
800	76.66%
900	78.97%
1,000	80.96%

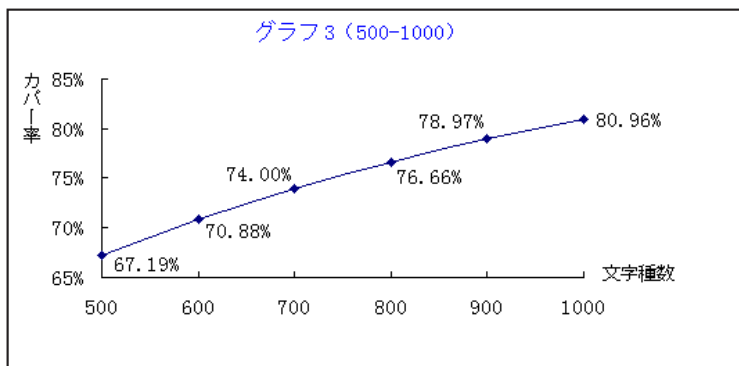
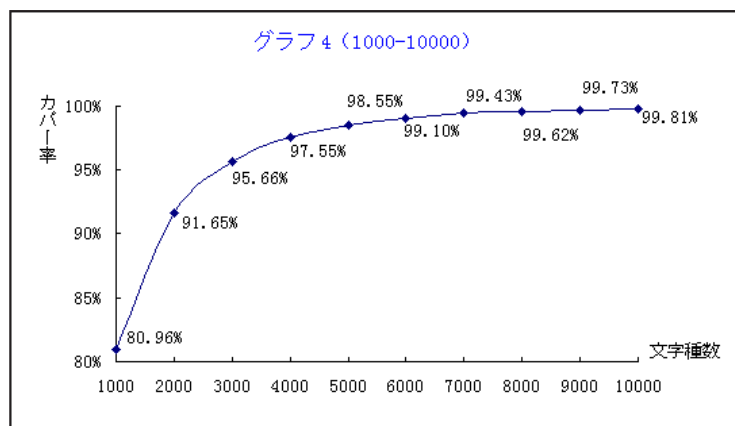
表 3

文字種数	カバー率
1,000	80.96%
2,000	91.65%
3,000	95.66%
4,000	97.55%
5,000	98.55%
6,000	99.10%
7,000	99.43%
8,000	99.62%
9,000	99.73%
10,000	99.81%

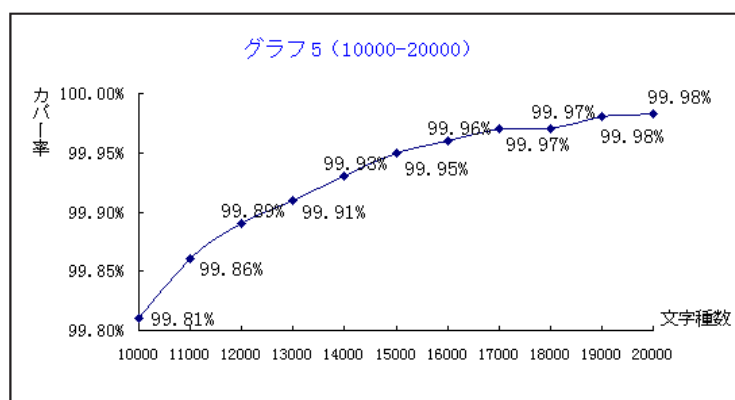
表4

文字種数	カバー率
10,000	99.81%
11,000	99.86%
12,000	99.89%
13,000	99.91%
14,000	99.93%
15,000	99.95%
16,000	99.96%
17,000	99.97%
18,000	99.97%
19,000	99.98%
20,000	99.98%

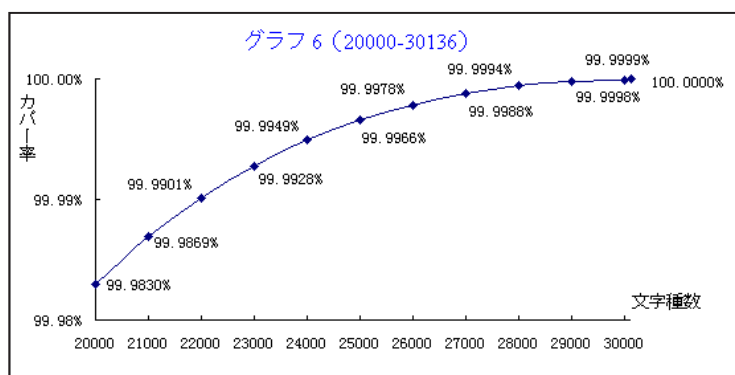
表5

文字種数	カバー率
20,000	99.9830%
21,000	99.9869%
22,000	99.9901%
23,000	99.9928%
24,000	99.9949%
25,000	99.9966%
26,000	99.9978%
27,000	99.9988%
28,000	99.9994%
29,000	99.9998%
30,000	99.9999%
30,135	100.0000%

表6

	使用文字充足率 9999‱	使用文字の完全充足	漢字 8000 字のカバー率	コーパスの使用漢字文字種総数	コーパス総文字数
『四庫』＋『四部』	総文字数：21,938 最後の文字：嘺 出現回数：23 出現頻度：0.0002‱ code：E411	総文字数：30,097 / 30,135 最後の文字：硒 / 碽 出現回数：1 出現頻度：0.0000‱ code：7852-78B5	カバー率：99.6186% 8000 番：龍 出現回数：1,165 出現頻度：0.0147‱ code：EA7C	30,135 最後の文字：碽	788,693,134 （4,000 部の著作、39,000 余冊）
『四庫』	総文字数：21,326 最後の文字：硶 出現回数：21 出現頻度：0.0003‱ code：7893	総文字数：29,049 / 29,081 最後の文字：帼 / 鎏 出現回数：1 出現頻度：0.0000‱ code：5E3C-93A5	カバー率：99.6794% 8000 番：鮀 出現回数：874 出現頻度：0.0125‱ code：9B80	29,081 最後の文字：鎏	698,076,596 （3,500 部の著作、36,000 冊）
『四部』	総文字数：22,410 最後の文字：䌑 出現回数：3 出現頻度：0.0003‱ code：4311	総文字数：27,602 / 27,606 最後の文字：蚜 / 蝾 出現回数：1 出現頻度：0.0000‱ code：6B5A-8767	カバー率：99.4879% 8000 番：㐌 出現回数：176 出現頻度：0.0194‱ code：EA83	27,606 最後の文字：蝾	90,616,538 （504 部の著作、3,134 冊）

表７　古籍漢字カバー率摘要

3. 古籍における漢字カバー率の分析

A. 中国古籍における通用漢字数の増大

ここでいう通用漢字数とは、文献をデジタル化する際に、要求を満たす文字数（すなわちカバー率）が 9,999‱（一万分の 9,999）に達する、すなわち欠字もしくは外字が 1‱（一万分の 1）以下になる漢字数を指す。

現代中国語における通用漢字数は、中国国家図書館が 1990 年代におこなった統計調査によると、約 7,000 字である。

今回の統計が示す中国古籍の通用漢字数は、約 21,000 字、すなわち文字出現頻度表の上位 21,000 字のカバー率がようやく 9,999‱に達する（表 7 参照）。この点は注目に値する。とりわけ、古典中国語字典を編纂する学者が文字を選定する際に重視すべきであろう。

古籍における通用漢字数が現代中国語の通用漢字数の二倍近くにおよぶ理由について、筆者は現代中国語では文字の規範化がすすんでおり、日常使われる漢字数が大幅に減少しているためであると考える。しかし古代においては、文字の規範がなく、異体字や書き癖による変体字などが多く、通用漢字数の増大を招いているのであろう。

B. 中国古籍の需要を満たす漢字数

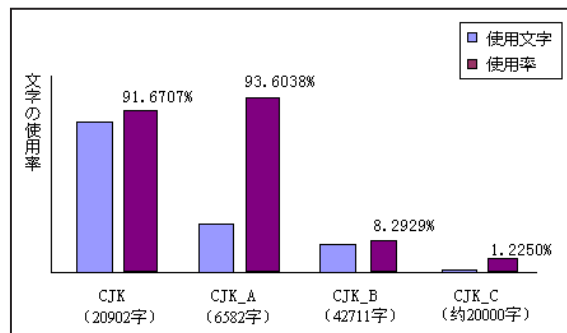
『康熙字典』などいくつかの大部の辞書を除けば、古籍における需要を十分に満たす漢字数は約 30,000 字である（表 7 参照）。『四庫』と『四部』の使用漢字数は、計 30,135 字であり、これは古籍における漢字使用状況を代表するものと言えよう。

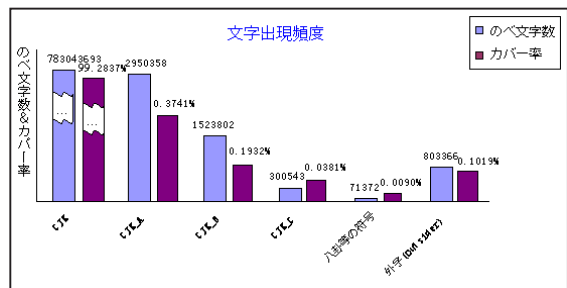
大部の辞書における用字を考慮するならば、50,000 字を突破するであろう。

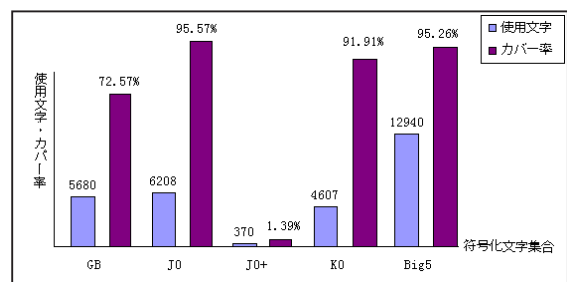
この統計は、ISO/IEC 10646/Unicode の CJK 統合漢字と CJK 統合漢字拡張領域 A の 27,484 文字の漢字集合が、中国古籍における漢字使用の需要をほぼ充たしていることをあらわしている。もしも私用領域の利用が可能であるならば、CJK 統合漢字拡張領域 B のうち使用頻度が高い約 5,000 字を外字として CJK 統合漢字と CJK 拡張領域 A に追加することで、Unicode の BMP の 2 バイト体系でも、十分に古籍通用漢字の処理を実現できる。

4. 漢字の各符号化文字集合における分布統計

A. ISO/IEC 10646/Unicode における分布

種類	定義文字数	使用文字数	使用率
CJK	20,902	19,161	91.6707%
CJK_A	6,582	6,161	93.6038%
CJK_B	42,711	3,542	8.2929%
CJK_C	約 20,000	245	1.2250%

B. CJK 被使用文字のカバー率

種類	のべ文字数	カバー率
CJK	783,043,693	99.2837%
CJK_A	2,950,358	0.3741%
CJK_B	1,523,802	0.1932%
CJK_C	300,543	0.0381%
八卦等の符号	71,372	0.0090%
外字（Outsider）	803,366	0.1019%

C. その他の符号化文字集合における漢字分布統計

符号化文字集合	使用文字数	カバー率
GB	5,680	72.57%
J0	6,208	95.57%
J0+[1]	370	1.39%
K0	4,607	91.91%
Big5	12,940	95.26%

5. 古籍漢字の各種符号化文字集合における分布

A. GB 2312 における分布

GB 2312 は漢字 6,763 字を定義するが、古籍（『四庫』・『四部』）で使用されるのはそのうちの 5,680 文字のみであり、カバー率も 72.57% に過ぎない。それらの漢字の大半は、簡略化されていない旧字体である。GB 2312 のカバー率が低いのは、GB 2312 が現在中国で用いられている簡体字文字体系のための符号化文字集合であり、2,200 字あまりの簡体字が定義される一方、古籍における使用は基本的に考慮していないためである。ここから、GB 2312 は現代文献の使用文字カバー率が 99.99% に達するとはいえ、古籍には適さないことがわかる。

B. JIS X 0208（J0）における分布

日本の国家規格である JIS X 0208 は、漢字約 6,500 文字を定義する。古籍（『四庫』・『四部』）では、そのうち 6,208 字の漢字が使用され、カバー率は 95.57% と比較的高い。これは、JIS X 0208 が収録する大多数の漢字が旧字体であるからである。日本独自の国字、約 280 字は、古籍においては基本的に使用されない。

C. KSC 5601 における分布

KSC 5601 は韓国の基本漢字集合で、漢字約 5,000 字を収録する。古籍（『四庫』・『四部』）ではそのうち漢字約 4,607 字が利用され、カバー率は 91.91% である。KSC 5601（K0）が収録する大多数の漢字は、旧字体である。

D. Big5 における分布

Big5 は台湾の常用符号化文字集合で、漢字 13,500 字あまりが収録される。古籍（『四庫』・『四部』）ではそのうち漢字 12,940 字が使用されており、カバー率は 95.26% に達する。カバー率は低くないが、しかし Big5 でも古籍における漢字使用の需要に十分こたえることができず、千字ごとに 50 字近くの欠落が生まれることを、統計は示している。Big5 は基本的には漢字常用文字集合に分類されるものである。

6. 中国古籍になぜ簡体字と常用漢字体が出現するのか

今回の統計では、きわめて低い確率ではあるが、古籍における簡体字と常用漢字体の使用が認められた。これらは主に『四部叢刊』に出現する。『四部』では編纂に際して、最も古い版本を収集することに特に力が注がれたが、その中には日本の正宗寺所蔵抄本を選んだ『春秋正義』のような例がある。日本の学者が書写するとき、しばしば「國」を「国」、「釋」を「釈」、「據」を「拠」などと表記する。『四部』にはまた後世の人物の手書きの序跋が収録されており、行書で書かれていると、たとえば言偏などが時として簡略化されている。従って、デジタル化に際して原書の姿を保存する原則をとったからには、簡体字や常用漢字体が出現するのも当然のことなのである。

訳注

[1] JIS X 0212

唐代ナリッジベースに向けて

Christian Wittern（クリスティアン　ウィッテルン）

訳：秋山　陽一郎（あきやま　よういちろう）

回 はじめに

数年前、デジタル化された膨大なデータが検索可能になったことは、中国研究の分野にまったく新しい方向性を与えた。今日､我々は『四庫全書』や、その他の多くのテキストデータベースを有し、そのうちの一部はインターネット上で無料で利用することができるが、それらのデータベースから受ける恩恵は限定的であった。これには様々な理由があり、そのすべてが技術的な理由によるわけではないが、技術的な面についていえば、それぞれのデータベースが個別に独立して提供しているインターフェイスと、それらによって個別的に運用される体系化されていないデータこそが、おそらく最も重要な要因となっているといえよう。

唐代ナリッジベースは、唐代に関する情報の包括的な電子アーカイブを、新しい手段によって検索・分析・拡張できるように提供することによって、こうした状況を改善しようという試みである。利用者にとってメインとなる検索手段はウェブベースのアプリケーションになるが、ほかのインターフェイスも開発が予定されている。

初期段階では、ほとんどの情報はテキストデータになるが、順次、画像や動画・デジタル化された地図などによってによって増強していく。ナリッジベース最大の特徴は、各情報が柔軟で斬新な方法によって結びつけられる点にある。

ナリッジベースにおける情報は以下のような情報軸に沿って構成される。

- 個人名、唐代の人々の事跡や履歴に関する情報
- 地名とその所在、行政区画の情報、デジタルマップ
- 文章・美術作品や建造物を含む、唐代につくられた作品に関する情報
- 暦と時間
- 重要事件とその影響

ナレッジベースにおける情報は、すべてではないにしても、これらの情報軸の複数にまたがるものが相当数あり、内部的に相互にリンクしつつ網目のような構造を形成している。加えて、これらの項目は階層的オントロジーを成していて、これにより階層内の位置や他の項目との相関関係からも情報を検索できるようになる。地理情報の場合なら、たとえば都市の情報は所属する上級の行政単位に関連づけられ、人物情報の場合なら、家系図のほか、出身地や、学統――僧侶の場合は授戒や嗣法の系統にも関連づけられる。

このような階層データは、情報の検索と抽出効果を高めるために、様々な形で活用できる。具体的な例を挙げておくと、従来の検索ツールでは、則天武后政権下でのし上がった官僚の概要や、唐王室復権後にその官僚たちがどうなったかを調査することは困難であっただろう。このような調査には何百人もの人物の伝記を広範囲にわたって研究し、個々の事跡を追った上でそれらの包括的な概観を把握することが必要とされる。これに対してナリッジベースでは、このような疑問に対してすべての関連情報を即座に集約し、しかもその結

果を解りやすく表示することも可能である。

歴史研究家の間でプロソポグラフィー研究として知られるこの種の疑問は、特にこのナリッジベースによって、家族関係・出身地・門閥情報を含む、唐代に活躍した広範かつ正確な個人データが完備した時点で公開されるだろう。

また白居易が任官中に出会ったすべての詩人や官僚たちを追跡して彼らの詩を分析することで、白居易が彼らからどのような影響を受けたかをみるという利用法も考えられる。この場合、この時代の詩をさらに白居易と直接面識のあった人物のものと、なかった人物のものとに大別して、その上でそれらの詩を比較分析することもできる。

もちろん、ナリッジベースが提供しうる活用法や課題はまだまだたくさんある。ちょうど一枚の紙に書かれる内容が何ものにも制約されることがないのと同様に、ナリッジベースの構成は、極力データに対して特定の解釈を押しつけないようにしようとするばかりか、むしろ相矛盾する観点の並存すら許容できるようにしている。

さらに技術的な視点からみると、ナリッジベースは以下の３つの柱で構成されている。

- TEI のタグセットを用いて、XML データベースにより検索可能な、唐代文化に関する原典テキスト群
- 事物・文章・発言、その他の多様な情報（いわゆるメタデータ）を管理するトピックマップ
- 原典テキストやメタデータ、その他すべての項目を結合し、ナリッジベースを運営するためのシステム

ナリッジベースの構築には、このようにコンテンツばかりでなく、適切なインターフェイスの構築も必要で、最善の結果を得るためには、この両方の要素が完全に統合されていなければならない。

回 原典テキスト群

原典テキスト群の開発は、このナリッジベースの心臓部にあたり、ほとんどの労力がこの作業に向けられることになるだろう。唐代に存在、もしくは制作されたすべてのテキストを幅広く提供することがこのナリッジベースの目標だが、一方でさらに急務とされているのが、唐代における人物や事件についての歴史的な史料からの情報を [1]、文字中心のテキスト検索よりも高度な形式で提供できるように準備することである。これは筆者が “deep markup” と呼ぶ、単にテキストの構造を示すという枠を超えて、テキストのフィーチャーにまで立ち入ったマークアップによって達成される。

このマークアップ作業によって、原典テキスト群は単なる符号化された文字列ではなく、それぞれのテキストによって与えられる内容情報に対しても、上記のように機械可読な形で利用できるようになる。たとえば、多くの読者は、テキスト中の「李世民」に関する記述が､同一人物である「唐太宗」と関連するという事をご存知だろうが、実際にはコンピュータではこれらはまったく別の文字列として処理されてしまう。「唐太宗」と入力した者と「李世民」と入力した者とが同じ情報を得られるように、これらの文字列の間の関係は明確にされなければならないのである。もちろん、このほかにも多様な人物や集団の相関形態があるが、テキスト中のすべての場所から、適切な人物について言及されている個所を検索可能にする意義は決して小さくない。このように、論理的には分かれていながら、実はナリッジベースのテキストレイヤー（原典テキスト群）と密接に関わっている情報レイヤー（メタデータ）には、人物や集団などの要素に対する識別情報を付与することが必要になる。一方でテキストレイヤー側でも、李世民に言及しているすべての個所に対して参照キーが付けられ、この参照キーによって情報レイヤーへの参照ができるようになる訳である。「太宗」とあるすべての場所にはもちろん同じ参照キーが付けられることになるが、ほかにも「文武聖皇帝」などは太宗と同等に扱われる。そのほか、たとえば「秦王」なども李世民ととてもよく関連するかもしれないが、同時に秦という国の王としても関連づけられるかもしれない。このような場

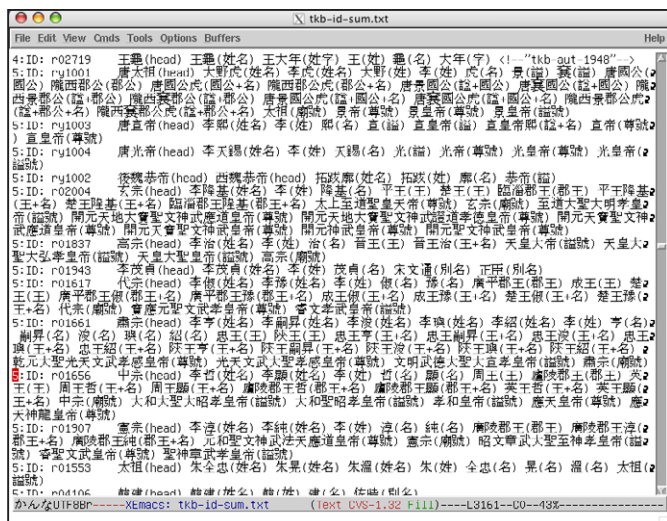
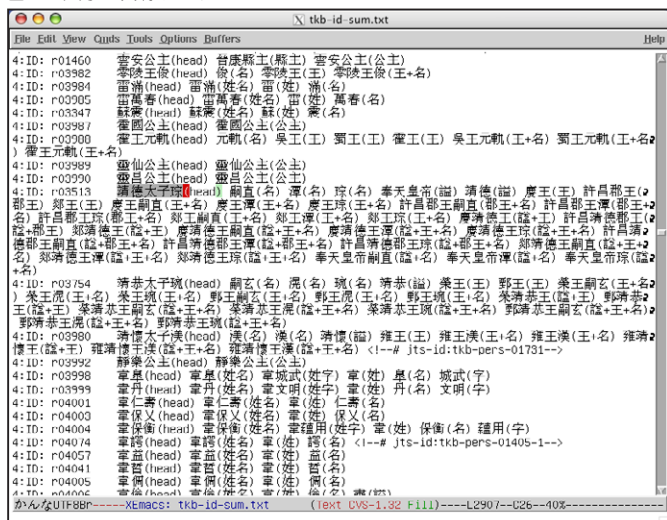



図 1　皇帝の異称リスト

合、有効な検索結果を出すためには、情報レイヤーとリンクするのに曖昧さを解消しておく必要がある。

“deep markup”の作業は、司馬光の『資治通鑑』（今日では目にすることができない失われた史料を豊富に用いた素晴らしい歴史書であると同時に、情報を時系列に並べ直しているという利点をも併せ持った歴史的作品）から手をつけることになる。この書はむしろ、事件や人物に関する機械可読な日付情報を抽出するのにそのまま利用できる。『資治通鑑』は、周の威烈王二十三年（前 403）から後周世宗の顕徳六年（959）までの期間を扱っているが、まず我々は 618 年から 907 年までをカバーする唐紀のみを利用する。正史である『新唐書』や『旧唐書』も重要な資料であり、もちろんこれも適宜活用していくことになるだろう。さいわい、その全文がすでに中央研究院のウェブサイトにて公開されているため、二重の労を強いる必要はない。『資治通鑑』や『新唐書』・『旧唐書』といったテキストを選択したのは、それらが大唐帝国の後半期、一世紀もの期間をかけて編集され、唐代に関する最も広範かつ精密な記事を提供しているからにほかならない。『資治通鑑』の内部構造はむしろ複雑になっている。我々は司馬光が用いた資料について、校訂・解説・考証・評価が加えられている胡三省の注釈本を利用しているが、胡三省の注にはさらに地理や官制に関する事柄や字音に関する注など、広範な注釈が施されている。『資治通鑑』を用いるのは、この胡三省の注釈本の構造を適切にマークアップして、その情報をトピックマップのために寄与させることがその狙いの一つである。テキストの内部構造に加えて、事件の起こった日時、人物の名前、地名、作品の題名や書物などもまたマークアップの対象となる。

その他のテキストについては、時間や需要と相談しながら作業を割り当てていくことになる。比較的少ない労力（高いコストパフォーマンス）で多大な情報をもたらすことができるマークアップ項目が優先的に追加されるだろう。百科事典としての類書などは、様々なテキストの中でも特に関心を寄せるべき対象といえる。

全体としては、ナリッジベースの

図 2　太子の異称リスト

開発スケジュールは、最初の二年間は全ての専門分野に関して下地となるべき情報を供給し、その後の三年間で宗教研究・学術史・科学史・美術史・文学史といったより専門的な分野のコンテンツを開発する予定である。

回 メタデータ

上述のように、ナリッジベースのメタデータはすべてトピックマップで管理される。トピックマップは、任意の情報とその関係を記述することができ、きわめて単純かつ抽象的な情報モデルに基づいている。この柔軟性と応用性は、メタデータ管理にとりわけ適しているといえる。メタデータの収集と処理は以下のように進められる。

- メタデータは可能なかぎり既存の電子化一次文献（『資治通鑑』中の固有名詞、または『旧唐書』『新唐書』の経籍・芸文志（書籍目録）、地理志など）から取得する
- このようにして取得された情報を処理・精錬しなくてはならない

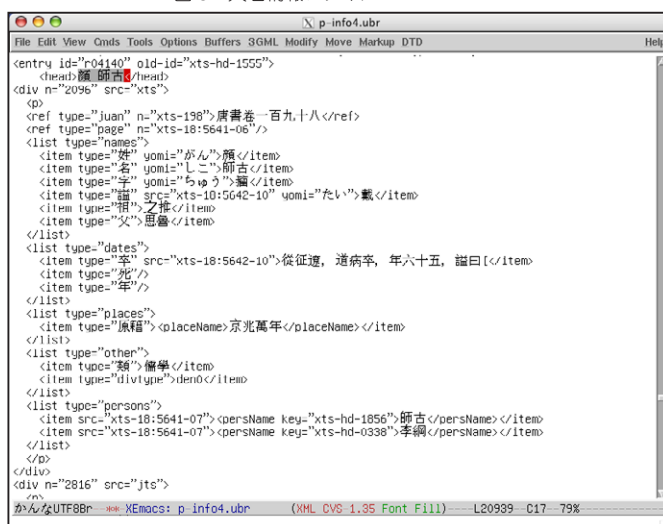
たとえばこれらのテキストから取得された固有名詞は、人名・地名の識別さえなされれば、より高度な情報を提供できるようになるだろう。地名は、位置や行政区画に関する情報を持った地理志やそれに近しい資料から得られた情報と重ねあわせることができるようになる。似たようなプロセスは、伝記中の重要事項と関連づけてトピックマップの形に変換することで、人名に対しても適用できる。

- 他の資料（紙媒体・電子化データの両方）から得られた近代以前や現代の情報もトピックマップにしかるべき手段で編入する。その際、すべての情報項目の出処も明示する
- 可能な限り、適用可能なすべての場所について地理的位置情報を追加する。これは地図をベースにした検索や、地名から地図上の場所の検索を実現するために利用される
- 日付や関連情報もマークアップ・正規化する。これは対照年表や日付からの検索などを実現するために利用される
- トピックマップ中のあらゆる情報は、原典テキストに含まれるより詳細な情報を探すのに利用することができ、また発見した情報をトピックマップに還元することもできる

回 ナリッジベース システム

ナリッジベースのシステムは、既成のコンポーネントやソフトウェア・アプリケーションをもとに開発され、主要な労力はそれらを結合させ、一貫した使いやすいインターフェースを提供することに注がれる。中核となる部分はオープンソースのウェブアプリケーション開発環境である Zope と Plone ベースとなる。テキストデータは、eXist といった dbXML 互換のデータベースに蓄積され、トピックマップ エンジンや編集用

図３　人名情報ファイル


```
p-info4.ubr
File Edit View Cmds Tools Options Buffers SGML Modify Move Markup DTD Help
<entry id="r04140" old-id="xts-hd-1555">
    <head>顔 師古</head>
<div n="2096" src="xts">
  <p>
  <ref type="juan" n="xts-198">唐書卷一百九十八</ref>
  <ref type="page" n="xts-18:5641-06"/>
  <list type="names">
    <item type="姓" yomi="がん">顔</item>
    <item type="名" yomi="しこ">師古</item>
    <item type="字" yomi="ちゅう">籀</item>
    <item type="諡" src="xts-18:5642-10" yomi="たい">戴</item>
    <item type="祖">之推</item>
    <item type="父">思魯</item>
  </list>
  <list type="dates">
    <item type="卒" src="xts-18:5642-10">從征遼, 道病卒, 年六十五, 諡曰[</item>
    <item type="死"/>
    <item type="年"/>
  </list>
  <list type="places">
    <item type="原籍"><placeName>京兆萬年</placeName></item>
  </list>
  <list type="other">
    <item type="類">儒學</item>
    <item type="divtype">den0</item>
  </list>
  <list type="persons">
    <item src="xts-18:5641-07"><persName key="xts-hd-1856">師古</persName></item>
    <item src="xts-18:5641-07"><persName key="xts-hd-0338">李綱</persName></item>
  </list>
  </p>
</div>
<div n="2816" src="jts">
かんなUTF8Br --**- XEmacs: p-info4.ubr      (XML CVS-1.35 Font Fill)----L20939--C17--79%
```

フレームワークとしては、現時点では Ontopia 社の Knowledge Suite の利用を考えている。

第一段階では、XML テキストベースに対するウェブベースの検索インターフェイスを開発する。これはその先のナリッジベース システム開発の根幹となる。次の段階では、トピックマップの情報と原典テキストとを重ね合わせて、トピックマップ エンジンを統合する。

回 期待される効果

唐代ナリッジベースによって、情報への新たなアクセス方法が実現する。検索可能な情報が限定されていたこれまでの文字列検索の代わって、知識空間を自由に航行することが可能になるだろう。これはちょうど遺跡の発掘のように、情報の大海の中からまったく新たな発見をもたらしてくれるはずだ。以下にナリッジベースによってできるようになることを幾つか例示しておこう。だがもちろん、可能性はこれらの例に尽きない。

歴史上の人物たちがどのように分布していたのかを、一定の高位を獲得した者と官吏とを異なる色で示しながら中国の地図上で表現することができる。その同じ情報を利用して、北方出身の人々と南方出身の人々とを対比させるなど、データベースの全部（もしくは一部。たとえば『全唐文』など）に対して絞り込みをかけることもできる。もちろん、詩の句末の文字に限定して検索をかけることもできる。このように利用者に対して、本文中で関心のある部分のみを的確に抽出するためにメタデータを重ねあわせる機能も実現してくれるのである。

ほかの用途では、高官を輩出している門閥の数やその出身地をたどってみるなど、高級官吏の社会的背景を比較してみるのも面白いだろう。

また、過去のある時期に、ある場所に在世していたすべての人物を特定するのも面白いだろう。たとえば仏牙舎利が長安を行進したのがいつで、それがどのような形で記述に反映されているかを、当時それを見聞きしていた人物の著作から明らかにするなど。

残念ながら、このナリッジベースが実現するには、長い時間と不断の共同作業を要し、近い将来にただちに達成されるものではない。当面の成果は、将来の開発のためのより深い考察に資するための試験環境用のプラットフォームの提供くらいになるだろう。同時に唐代ナリッジベースに興味を持つ学生のために、有益なサービスを提供することが予定されている。

注

[1] すでに幅広い原典のテキストデータを収めた中国歴代基本典籍庫 隋唐五代巻 CD-ROM が 国学から発行されており、テキストデータそのものに注力すること自体にはそれほどの意義はない。

漢情研 2004 年春期・夏期公開講座報告

デジタルアーカイブと東洋学 東洋学と著作権 II

本研究会では、昨年度より資料をデジタル化する際の様々な問題点について理解を深めるべく、公開講座を開催している。

本年度は、春期公開講座（2004 年 3 月 20 日　花園大学にて開催）において中国学最大のテキストコーパスである「台湾中央研究院漢籍電子文献」の開発に関わった陳弱水氏をお招きしてお話を伺った。また夏期公開講座（2004 年 7 月 19 日　早稲田大学にて開催）では、法学者の石岡克俊氏を昨年に引き続いてお招きし、更に本年は丸善の笹井真也氏に実際に資料のデジタル化に携わった経験を踏まえてのお話を伺った。

以下は、各公開講座の講師による論攷、および報告である。

※漢情研 2004 年春期・夏期公開講座の日程・会場等の詳細については、彙報（p.180）を参照していただきたい。

Contents

デジタルアーカイブと東洋学

中央研究院歴史語言研究所の経験から

陳 弱 水（ちん　じゃくすい）
訳：山下　一夫（やました　かずお）

◇ はじめに

本日は漢字文献情報処理研究会で発表の機会を頂き、大変光栄に思っております。会の方からは「デジタルアーカイブと東洋学」というテーマを頂いておりますが、私ができる範囲で中央研究院歴史語言研究所（以下、中研院史語所と略）における漢籍データベースの状況についてお話をさせて頂きたいと思います。

ただ発表をするにあたって不安に思うことが三点ほどあります。まず第一点。私が史語所の漢籍データベース計画を担当したのは1994年から1997年までの間で、それからもう何年も経っていてこの方面の問題にはだいぶ疎くなっています。そのため、私の発表もだいぶいいかげんなものになってしまうのではないかと心配しております。次に第二点。私は昨年の9月末に来日いたしましたが、資料をほとんどもってきておりません。今回お話をさせていただくにあたって、史語所の方に幾つかの資料を送ってはもらいましたが、それでもまだ決して充分なものとはいえません。そして第三点。以前、貴会の『漢字文献情報処理研究』第二号に、漢籍全文データベース計画に関する私の原稿の翻訳を掲載して頂きました。その原稿は2000年の夏に執筆したもので、それから3年半しか経っておらず、状況にあまり大きな変化はありません。そのため、私の発表も新味に欠けるものになってしまうのではないかと心配しています。

本日発表させて頂くにあたり、こうした不利な条件がある以上は、一体どういった角度から史語所の漢籍データベース計画を紹介すれば、ご招待くださった皆さんのお気持ちに沿えるものかと、色々考えました。そして、あるいは「インサイダー」としての視点から、私の知っていることや考え方について、ざっくばらんにお話しすることで、皆さんに台湾の、特に中研院史語所における漢籍データベース開発の状況について、よく理解していただけるのではないかと考えた次第です。

私が漢籍データベース計画の責任者をつとめたのは3年間だけでしたが、文章で、あるいは口頭で、外部の方々に対して史語所のこうした仕事を紹介するという機会が少なからずありました。それは、私が担当した期間がちょうど漢籍データベースの重要性が国際的に重視され始めた時期だったからです。私が在任中の1997年に、「二十五史」データベースの「表」の部分が完成して「二十五史」計画が正式に完了し、さらにこれ以外の事業を積極的に展開していました。そのため外部の方々が私たちに接触してくることが次第に多くなり、そうした環境の中で私は対外的に我々の仕事について説明する義務を負っていました。その後データベースの仕事からは離れました

が、行政的な職務も担当していたせいで、データベースについて研究所を代表して発表することもありました。そのほとんどは学会での正式な発表で、主に私たちの作業の成果についてお話しするというものでしたが、本日は角度を変えて、学術界における同業者という立場から、事業内容と内部の状況についてご紹介したいと思います。

本日お話ししようと思っておりますのは、以下のような問題です。第一に、2000年の夏から現在までの史語所の漢籍データベース計画の発展の状況についてです。あわせて、史語所のその他の文献資料電子化についてもご紹介したいと思います。第二に、中研院と史語所におけるデータベース計画の運営・事務上の問題についてお話ししたいと思います。第三に、データベース開発事業についての私個人の回顧です。

◇ 史語所のデータベース

◉ 漢籍全文データベースの歴史

さて、中研院が中国古典文献データベースを開発した背景と歴史については、私やその他の中研院のメンバーが、これまでに色々なところで紹介してきましたので、くり返して説明する必要はないかとも思いますが、皆さんに最低限の情報を共有して頂くために、簡単にご説明したいと思います。

中研院が漢籍データベースの作成を始めたのは、1984年初めのことです。史語所・情報科学研究所・計算中心（コンピュータセンター）の三つの組織が共同で打ち出した正史食貨志自動化の実験的プロジェクトが、台湾の行政院文化建設委員会（他の国では文部科学省に相当します）の資金援助を得て開始されました。このプロジェクトが一年後に一応の成功を収めたことで、二十五史全文データベースの作成が正式に始まりました。この過程におけるもっとも重要な成果は、「中文文献処理システム」ソフトを開発したことで、これは後に正式に「瀚典全文検索システム」と命名されます。

以上の簡単な説明からもお解りになると思いますが、漢籍データベース計画（当時は「史籍自動化」と呼んでおりました）は、全くの独創的事業であり、ソフトもなく、前例となる作業工程もない状況の下で始められたのです。この画期的な事業には、大変多くの方々が関与しましたが、キーパーソンは毛漢光氏と謝清俊氏であったと、私は理解しています。毛氏は著名な中国中古史研究者で、この計画を立案し、かつ積極的に推進された方です。また謝氏は情報科学の専門家で、もともと漢字入力について長年研究されており、かつまた人文学についての大変な素養をお持ちです。彼こそが、中研院の情報処理技術を史籍自動化計画に持ち込んだ重要人物です。私がここで重ねてこうした昔話めいたことを持ち出したのは、中研院史語所の漢籍データベースの開発が先駆的たり得た決定的な理由は、人文学者と情報科学の専門家との協力にあったということを、強調しておきたいからであります。こうした条件は、他の多くの機関では90年代半ばになっても実現できませんでした。なお、私が史語所に着任したのは1993年ですので、以上はいろいろな文献資料や、同僚たちの回想に基づいたものではありますが、これで間違いないはずです。

それでは、2000年の夏から2003年の終わりまでの3年半の間における、漢籍全文データベース計画の作業状況のお話に移りましょう。この期間、データベースの規模は引き続き安定的に成長しました。2000年夏の段階では、漢籍データベースでオンライン公開されたデータ完成部分の総文字数は約一億四千万字でしたが[1]、昨年末の段階では約三億字にも達しており、増加分は一億六千万字と、実に倍以上の規模に成長しているのです。詳細については付録をご参照ください。

この三年半の間に増加したのはどういったデータベースでしょうか。まず史部については、最も大きいは『清実録』で、そのほかにも幾つかの個別の史料文献が増えておりますが、それらは主として宋代と元代のものです。『清実録』は四千万字以上あり、これだけで二十五史の分量を超えております。経部と子部についてはあまり変化があ

りません。仏典のデータベース化はすでに停止しています。これは「中華電子仏典」（CBETA）の作業との重複を避けるためです。道教経典は『抱朴子』が加わっただけです。医薬文献については約二千万字が増えておりますが、これは史語所で医療史研究が活発に行われていることを反映しています。文集と筆記小説は、約一億字と最も増加しております。最も重要なのは、『全上古三代秦漢六朝文』・『全唐文』・『全宋詩』・『全遼文』・『国朝文類』などの総集です。このほか、歴代史料筆記が全部で約千八百万字増加しました。最後に、政書、類書部分は約一千万字増加しております。

ここ三年あまりの努力によって、漢籍全文データベースは目下、宋五代以前の現存文献のデータベース化をほぼ完了しています。史語所以外では、『全唐詩』と『太平広記』もネット上で検索できるようになっています[2]。先秦から唐代までの歴史文献研究については、基礎的な電子ツールが概ね完成したといって良いでしょう。宋元代のデータベースも続々と増加中です。明清の資料は膨大な量にのぼるため、データベース化の比率はまだ非常に低いですが、『明実録』の完成が明代の歴史文献研究の大きな助けとなることは疑いありません。史語所が現在開発中のデータベースは約四千五百万字で、ほとんどは宋元明の資料ですが、『冊府元亀』・『文苑英華』・『北堂書鈔』、あるいは古代や六朝の小説など、多少これより時代の古いものも含まれています。このほかにもデータベース作成が決まっている文献資料が多数あり、その中には『正統道蔵』の一部も含まれております。

⦿ その他の歴史文献データベース

つづいて、史語所で進められているその他の歴史文献の電子化事業についてご紹介したいと思います。まず初めにご説明したいのは、史語所の漢籍データベース計画が成功を収めた後、中研院内のほかの人文関係の研究所の多くが、史語所の作業ノウハウを借りて、それぞれのデータベース計画を打ち出したことです。その中でもっとも成功を収めたのは台湾史研究所で、ここは膨大な台湾史データベースを作成しました。また言語学研究所も、私たちが入力した資料を利用して、言語学者の利用に適したデータベースを作成しました。しかし、その他のあらゆる研究所でデータベース化が順調に進んでいるわけではありません。そうした状況から考えるに、史語所が漢籍データベース作成に成功したのは、まことにスタッフの長年にわたる努力の賜物なのです。もしワーキンググループを組織し、経費を調達しようという強い意志が無ければ、電子化事業を進めることは決して容易ではないのです。

実際にはここ数年、史語所内部の、あるいは史語所と関係がある電子化計画は非常に多く、現在では漢籍全文データベース計画は私たちの電子化事業の一部分にすぎません。最重点課題ではないといってもいいでしょう。漢籍全文データベース計画が成功したことで、史語所のスタッフの多くがデータベースの作成に興味を持つようになり、彼らの努力によって漢籍全文データベースとは異なるタイプのデータベース計画が立ち上がったのです。これらの事業は様々な方式で進められていますが、スタッフによる自発的なものがほとんどであるという特徴があります。研究所内の特定の研究室が進めているものもあれば、スタッフが個人で進めているものもあり、また他の機関の研究者と協力して行っているものもあれば、また中研院内の他の組織と協同製作しているものもあります。これらの計画は経費の出所も複雑ですが、大多数は自主的に外部資金を申請しています。史語所の現在のデータベース事業の概況については、史語所サイトの「学術資源」以下の「資料庫」の項目をご覧ください（http://www.ihp.sinica.edu.tw/index_page.htm）。

ここでは、これらの計画のうち主なものを幾つか紹介したいと思います。研究所内のプロジェクトとして作成されたデータベースには、代表的なものが二つあります。一つは「文物図像研究室資料検索系統」です[3]。これは、「簡帛金石資料書目」・「居延漢簡補編図像系統」・「武氏祠画像検索系統」・「安丘董家荘漢墓画像」などが含まれます。これは漢代を研究しているスタッフが中心となり、研究所内の文物図像研究室での作業を通じてこつ

こつと作成したものです。このデータベースの特色は、文字検索を通じて、関連する画像にジャンプできるという点です。もう一つ紹介したいのは、内閣大庫檔案（訳注：檔案とは公文書・保存文書のこと）のデータベースで、これは研究所サイドがサポートする公式プロジェクトになっています。史語所は大量の明清内閣大庫檔案（主に清代のもの）を所蔵しています。これは、研究所の移転や資金的問題のために、80年代になってようやく全面的な整理作業が始まったものです。90年代半ばに、整理作業に情報技術が導入されるとともに、檔案の基本的資料と概要を主な内容とするデータベースの作成も始まりました。檔案を利用する研究者が、データベースの検索によって簡単に必要な檔案を探し出し、またネット上から檔案の影印をプリントアウトすることができるように設計されています（これは目下の所、台湾国内での有料使用に限定されています）。この外にも、研究所のスタッフが、個人やグループで作成した書目データベースも幾つかサイト上にあがっており、外部に公開されています。要するに、史語所の「資料庫（データベース）」のページには、「漢籍全文データベース」以外にも、幾つかの小型のデータベースがあり、そのうち比較的規模が大きいのが、石刻画像資料（特に漢代と関係するもの）と、明清内閣檔案データベースである、ということです。ただこれらの多くは、研究所のスタッフが個人的に作成した資料を「寄贈」したものであるため、内容はあまり系統だっておらず、漏れや誤りも多く、正式な研究ツールといえるものではありませんが、こうした点につきましては外部の皆さまにご理解を頂ければ幸いです。

このほか、現在のところ史語所のサイトには明確に示されていない、二つのデータベース計画についても触れておくべきかと思います。一つは「中華文明之時空基礎架構（『中華文明の時空』の基本構造）」と称するものです。この計画は、史語所の研究員である范毅軍氏が中心となって中研院コンピュータセンター等と協力で作成しているもので、現在、中研院のサイトの「学術資源」の下に置かれています。（http://ccts.sinica.edu.tw/index.php?lang=zh-tw）。このデータベースは、基本的には地理情報データベース（GIS）です。すでに譚其驤氏主編の『中国歴史地図集』の電子化に成功しており、作成されたデータベースは、他のテキストデータベース（『二十五史データベース』など）とリンクしています。この計画は始まって数年になり、さらに発展し成熟した暁には、非常に有用なデータベースおよび研究ツールになることでしょう。

このほか「金文デジタルデータベース」もご紹介するに値しましょう。これは史語所の古文字および中国古代史の研究員と、他の研究機関の研究者との合同計画で、開始されてからすでに二・三年が経っております。その名の通り、古文字（特に金文）のデータベースの構築を目指しています。計画が完成すれば中国古代文明研究に大いに貢献することは間違いないでしょう。

以上二種類の他に、別のタイプの更に新しいデータベースもあります。行政院国家科学委員会（国科会）の「国家デジタルアーカイブ計画（National Digital Archives Program)」プロジェクトに属する事業です。ここ数年、科学技術と学術研究を主管する政府機関、国科会が大量の経費を投入し、「国家型計画（国家的プロジェクト)」というものを創設しました。これは、重要な研究に対して、重点的サポートという方法で資金援助してゆくというものです。この「国家的プロジェクト」の中で、唯一人文社会領域に属するのがこの「国家デジタルアーカイブ計画」です。2001年の初めに正式に始動しました。これは、決して独創的な研究プロジェクトではなく、名前の通り国内の（文字資料、非文字資料を含む）様々な所蔵品をデジタル方式で保存し、表現し、広めることを目的としています。これは統一規格をもつプロジェクトではなく、マスタープランで大きな方向だけを規定し、実際に執行されるプロジェクト（「子プロジェクト」）は、学術界・文化界から公募し、審査によって、どれを推進するか決定するというものです。この「国家デジタルアーカイブ計画」では中研院が重要な役割を担っており、その中でも、電子化事業に関する経験と実績を有す

る史語所が深く関わることになりました。

また、史語所が進めているプロジェクトには、史語所サイドが提案したものもあれば、研究員が個人で提案したものもあります。前に触れた研究員が個人的に作成したデータベースの中には、この「国家デジタルアーカイブ計画」の資金補助を受けているものもありますが、経費のすべてをこの計画から取得している訳ではありません。この辺の状況はやや複雑なのですが、要するに、史語所が進めている「国家デジタルアーカイブ計画」の子プロジェクトには、前述の研究員が自発的に進めているプロジェクトと重なっているものもあれば、重なっていないものもあるということです。史語所における「国家デジタルアーカイブ計画」の進捗状況については、史語所サイトの「学術資源」以下の「数位典蔵（電子所蔵品）」の項をご参照ください（http://saturn.ihp.sinica.edu.tw/%7Edahcr/）。これらのプロジェクトの中心は、史語所の所蔵資料の電子化です。例えば、考古出土品やそれに関連する写真や記録・檔案、石刻の拓本や古文書、傅斯年図書館の善本書、民俗学の調査標本や写真・檔案などが挙げられます。これらのプロジェクトは現在積極的に進められているところで、大部分はまだ公開できるような成果は上げておりませんが、傅斯年図書館の善本書のスキャンデータはすでに相当量蓄積されていますし、こうした作業の成果によって、将来的に研究資料の流通が非常に便利になるものと思われます。

このほか、史語所所蔵の敦煌文書は、数は少ないながら幾つかの重要なものがあります。私たちはすでに英国図書館（British Library）の「国際敦煌計画（International Dunhuang Project）」に参加しており、それらの文書はまもなく IDP のサイト[4]に公開され、全世界の研究者の使用に供されることになっています。

◆ データベース運営の実態

◉ 漢籍全文データベースの運営組織

史語所で進められている電子化事業の概要については これくらいにして、次に、史語所が歴史文献の電子化を進める上での運営方針という問題についてお話ししてみたいと思います。

中研院と史語所の漢籍データベースの運営方針をご理解いただくには、まずこの事業の組織方法についてご説明しなければなりません。先ほど史語所における歴史文献電子化の歴史についてご紹介した際に、この事業は主に史語所、あるいは研究員個人の立案によるものだということに、お気づきになった方もいらっしゃるではないかと思います。こうした状況は、後に変わりはいたしましたが、このような背景が、データベースの継続的な開発や管理方法といった問題に大きく影響しています。話が複雑にならないように、ここでは漢籍全文データベース計画の例をご説明しようと思います。

漢籍データベースは、当初、研究員が自発的に立案したものが、中研院内部のいくつかの機関の支持を得て、史語所ではじめられたものです。経費は外部から獲得したもので、行政院文化建設委員会以外に、国家図書館やアメリカのワシントン大学東アジア図書館などの援助を受けました。1989 年からは正規の予算もつきましたが、90 年代の中後期までデータベース開発経費の大部分は史語所が負担していました。史語所は漢籍データベースを開発したために、大きな負担を負ってしまったのです。これが、私が先ほど「史語所が漢籍データベース作成に成功したのは、まことにスタッフの長年にわたる努力の賜物なのです」と申し上げた理由です。ここ数年で、漢籍全文データベース計画の経費の大部分は中研院が直接負担するようになりました。

また、作業組織にも変化があります。古典の電子化の重要性が日増しに顕著になってきたことに鑑み、中研院は 1997 年初めに全研究院的な組織である「漢籍電子文献協調委員会」を発足させ、各研究所の電子データベース構築を調整・サポートすることになりました。2000 年の初めに「国家デジタルアーカイブ計画」が始まると、中研院は「中研院デジタルアーカイブ諮問委員会」を新たに設けて、所蔵品の電子化と既存のデータベー

ス構築プロジェクトを統括させ、従来の委員会を廃止しました。

現在、中研院では、デジタルアーカイブ（データベースも含む）に関わる意志決定は、いずれも二つのレベルで進められています。一つは中研院の委員会で、もう一つは作業を行う各研究所あるいは研究所内の委員会や研究員のグループです。

⦿ データ構築の方針

漢籍データベースの運営については、データの構築、データベースの公開と管理という、二つの側面に分けられると思います。データの構築については、史語所レベルでも中研院レベルでも、漢籍データベースを長期的に発展させてゆくというコンセンサスが得られています。「漢籍全文データベース」という名称から解るように、史語所が構築しようとしているのは、近代以前の中国を研究する学者の使用に供する総合的なデータベースであり、あらゆる伝統的漢籍文献が開発の視野に入っています。別の機関や個人が一定のレベルを有するデータベースをすでに開発していたり、あるいは開発を計画していたりする場合は、重複を避けるようにしています。例えば、仏典はすでに私たちの事業の範囲には入っておりません。

現在、史語所の漢籍全文データベース以外の、同じような性格の大規模データベースとしては、『四庫全書』と『四部叢刊』があります。この二つのデータベースと史語所の開発方針の違いは、私たちの対象は資料そのものであって、特定の図書ではないという点にあります。私たちは学術研究の視点からまず何を電子化するかを決め、それから版本を探すという方法を採っています。こうした方法は、『四庫全書』タイプのデータベースと機能面で相互補完しうるものですが、場合によっては私たちのやり方の方が効果的でしょう。たとえば、六朝史を研究するならば、私たちの『二十五史』・『全六朝文』・『通典』・『先秦漢魏晋南北朝詩』などのデータベースを使った方が、効率が良く、遺漏も少ないはずです。

現在、多くの国々で歴史文献のデータベース構築が進められています。アルファベットで表記される言語の場合は、データベース構築が比較的容易です。私の知る限りでは、フランスとイギリスでは近代に発行された出版物の大部分がすでにデータベース化されています。例えば、イギリスのオンライン図書館である Eighteenth Century Collections Online（ECCO）[5] は、1701 年から 1800 年にかけてイギリスで出版されたあらゆる出版物を網羅しており、約十五万種、三千三百万ページ以上にものぼります。作品は英語のものが大部分ですが、ラテン語・フランス語・ドイツ語・イタリア語およびその他のヨーロッパ言語も含まれています。イギリスでは民間の企業が歴史データベースの構築を行ったため、非常に高い使用料を支払わなくてはなりません。フランスのデータベースは政府のサポートを得ており、完全に無料で使うことができます。東アジアの国々の場合は文字の特性のために、歴史文献データベース構築は欧州よりも困難です。しかし史語所は、近代以前の文献についての完全なデータベースを学術界の協力のもとに作成することを長期的な目標としており、西洋の国々の水準にいずれは追いつきたいと考えております。

⦿ データベースの公開と利用

次にデータベースの利用に関する問題についてですが、現在は、大変有用な二十五史と十三経のデータベースを含む一部分を無料で公開しています。中研院内部では、研究員および正式な訪問者の全員が、全データベースを使うことができます。外部から未公開データベースを利用したい場合は、国内向けには有償接続という方式を取っており、個人でも機関でも可能です。国外については、状況に応じて、共同開発・文献交換・開発資金援助などの方法によって、学術機関に供用しています。

現在のところ、漢籍全文データベースは部分的公開という運用方針を取っております。これは難しい問題です。史語所は営利組織ではなく、データベース構築の長期的目的は公共の使用に供することにあります。しかし一方では、このデータベースの構築は、中研院と史語所が長年にわたって自分たちで費用を負担して進めてきたものであり、

そしてその費用はすべて台湾の納税者から頂いたものです。そのため、目下のところデータベースの使用は中研院内部と台湾国内を優先すべきという意見が多数を占めています。そのような状況であっても、ネット上で公開されている中国古典文献データベースは、史語所のものが最大です。今後、漢籍データベースの公開はいっそう進展することしかありえませんが、どのくらいのスピードで進むかについては、まだ簡単には判断できません。ともかく、東洋学研究者が自由に史語所のデータベースを活用する機会は将来的にもっと増えるはずです。

◆ データベースをめぐる問題点

◉ 運営組織の問題

次に、史語所のデータベース開発事業にまつわる私個人のいくつかの思い出をお話ししようと思います。史語所漢籍全文データベース計画の最大の長所は、着手された時期が早く、作業が持続して行われ、しかも専門の研究者が直接担当したことで、校正が厳密で本文の信頼性が高い、学術的利用価値の高い大規模データベースを製作することができたという点にあります。こうした長所については皆さんよくご存じのことと思います。ここでは、私が目にしてきました、幾つかの問題点についてお話したいと思います。

まずは、組織についてです。中研院のデータベース開発やそのほかの電子化計画は、ほぼ完全に学術研究の枠組みの中で進められたもので、これを専門に行う独立した組織が存在しません。しかし、電子化事業と学術研究という二つの事業は時に対立するものです。電子化計画の目的は研究ツールの製作や、文献資料を保存・利用・普及に便利な形式に改造することにあります。しかし中研院は独創的な研究を強調しています。一般的に言って、研究ツールの開発は正規の研究成果とは見なされないので、電子化計画の担当者は往々にして仕事に専念することができません。しかも、電子化を専門的に扱う組織が無く研究者の自由にまかされるため、結果作業が重複してしまうことがあり、また経験を蓄積してゆくこともできないのです。

◉ スタッフの問題

次に、研究者は計画の立案や、新しい電子化事業の開拓には優れています。しかし、一応の成果が出てしまうと、出来上がったものをさらに改良したり、きちんと管理したりということに興味を持たないことが多いのです。そうした仕事は事務スタッフに委ねられることになりますが、彼らは必ずしもデータベースやデジタルアーカイブの内容を理解しているとは限らず、容易に間違いを引きおこします。端的に言えば、営利企業に比べ、中研院のような機関がデータベースを開発する場合の弱点というのは、製品の整理や管理・普及という方面に傾ける労働力や意識が欠けているのです。

◉ データベースの改良の問題

次に、史語所漢籍データベースの改良の問題についてお話ししたいと思います。史語所の漢籍データベースは二十年前に構築がはじまったもので、設計や技術の面ではいささか古くなっています。データベースを改良する計画は無いのかと訊かれたら、答えはある、です。事実、長年にわたりIT技術の進歩とソフト環境の変化に対応するために、データベースにもそれ相応の調整を加えています。初期には、字が無いという問題が大変深刻でしたが、この問題はすでにかなり改善されました。ただ、人的・財政的の限界に加え、研究機構の特質が新たな事物の創造という点にあるため、大々的な改良を行うのが難しい状況にあります。

現在検討されている改良計画もあります。たとえば、データが原本の第何行目にあたるかを表示できれば（現在は頁番号しか表示されません）、検索の効果も上がり、研究者が元の文献に当たるのもより便利になるでしょう。また、改良計画の中で比較的重要なアイデアは、「典拠ファイル（authority files）」を増やすことによって、データベースが検索機能ばかりでなく、特定のタイプ

の基本データ（人名、地名、年代など）をも提供できるようにするというものです。ただ、個人的な意見では、学術的なデータベースの典拠ファイルをきちんと作成し、データベースにワークステーション的な機能を持たせるというのは、決して簡単なことではありません。典拠ファイルは、良質な工具書に基づいて作成されてはじめて信頼のおけるものになります。しかし、工具書を複製し電子化をする権利を取得するのは非常に困難であり、何度も折衝を重ねて行く必要がありますが、これは決して研究機関の得意とする分野ではありません。

中研院と史語所が80年代中期に電子データベースの開発を始めたのは、多かれ少なかれ偶発的な事件でした。この事件のために台湾は中国古典データベースの構築において、主導的地位を占めることになったのです。そして今に至るまで、ネット上の無料で使用できる主な漢籍データベースの大半は、台湾が提供しています。漢籍データベースが開発される以前、台湾における中国歴史文献研究は発達してはいましたが、しかし一般に辞典や工具書の編纂を重視しなかったため、その方面での貢献は微々たるものでした。大規模なデータベースの開発は、もとより台湾の学術界が誇るべき業績でありますが、別の面では、これは台湾の国際学術界に対するフィードバックであるともいえるでしょう。私たち台湾の学者は、長年、中国大陸で校定された古籍を使い、日本と大陸の学者が編集した工具書を使ってきましたが、今度は東洋学や中国歴史文献研究のインフラ建設に貢献をするとともに、実用主義が幅を利かせた時代、変化の速い情報技術の時代を、この学問分野が戦い抜いていくための力を与えることができたのです。

私のもう一つの意見は、電子データベースは決して孤立した研究ツールではなく、良質な研究用データベースを作るには、時として異なるタイプの研究ツールに深く依存しなくてはならないということです。例を挙げれば、史語所は唐代以前の文献のデータベース化をほぼ完成させており、作業の重点は宋元明に移っております。最も効率的に宋代の歴史文献データベースを構築するなら、『全宋文』と『続資治通鑑長編』のデータベース化を最優先で検討するべきです。これには学術界間の協力という問題が関わってきます。もし簡単に四川大学の『全宋文』と北京中華書局の『続資治通鑑長編』の電子化の権利を取得することができるのならば、漢籍データベースを順調に構築できることは間違いありません。しかし、現在の台湾海峡両岸の政治関係のもとでは、何事も簡単には済みません。また、私の考えでは、漢籍データベースに典拠ファイルを追加する便利かつ妥当な方法は、既存の様々な索引や辞書などを利用することです。つまり私が強調したいのは、電子データベースは見た目は新しいものではありますが、実際には伝統的な文献学や工具書とたいへん密接な関係を持っており、ある意味、既存の業績をさらに発展させたものであるということです。こうした関係をさらに深化させて、より質が高く便利な研究環境を作り出してゆくべきだと思います。

◇ おわりに

最後に、電子化事業と東洋学研究の関係についての、私個人の基本的な考え方をお話ししたいと思います。私がこの問題についてお話しするのは、貴会から頂いたテーマが、この問題を暗示しているように思えたからです。デジタルアーカイブと東洋学研究の関係については、多くの研究者が独自の感想と見方をお持ちでしょうから、ここでは一つだけ、特に珍しくもない平凡な意見をお話しいたします。

お話ししたいのは、あらゆる要素から考えて、伝統的中国歴史文献研究についていえば、電子化は必然であり、また電子化事業の中で、データベース――とりわけ良好な研究機能を提供できるデータベース――の開発は、特に重要であるということです。こうした作業は一般社会におけるインフラ整備のようなもので、必ず進めなければなりません。作るべきか否かという選択の余地はありません。問題は、どうすれば上手く作れるかということであり、それにもまして重要なのは、ど

う利用したら実質的なメリットを生み出せるかということです。デジタルアーカイブの効能は、作業効率を高め、以前の学者にはできなかったことを行い、さらには以前の学者が夢想だにしなかったことすらも実現する点にあります。しかしいずれにせよ、学術研究について言えば、電子化の成果がどんなに重要だといっても、基本的には依然としてツールにすぎません。便利なツールを通じて知識と理解を深めることこそが、学術コミュニティーの本当の目標であると私は思います。

特殊なめぐりあわせで、中研院史語所は率先して漢籍データベースを開発し、それなりの成果を上げました。中研院側の経費面でのサポートも得て、事業を維持することもできました。こうした状況の下で、学術界や実業界のほかの電子化事業とも協力して、より完全で大規模な中国古典文献データベースを作成することが、今や社会や学術界に対する史語所の責任となった感があります。研究所内の大部分のスタッフも、史語所が継続してこうした役割を担っていくことを望んでいるのではないかと思います。専門的な研究機関であるため、電子化計画に投入できる技術レベルおよび管理レベルの資源には限界がありますが、より便利で正確なツールの実現に向けて、力を尽くさねばなりません。データベースの構築は交通機関の建設とは違って、環境破壊や騒音公害は発生せず、直接的な欠点もありません。史語所について言えば、比較的大きな悩みは資源の分配です。つまり、研究機構として、こうしたインフラ建設事業にどの程度の財力と人員を充てれば、最も重要な本務に支障をきたさずに済むか、という問題です。

また、デジタル研究ツールが普遍化するということは、研究に良からぬ副作用をもたらすこともあり得ます。少なくとも私個人の台湾での経験からいうと、個人的に最も心配なのは、これのために大量の粗雑な研究が生まれてしまっていることです。それらはデジタル検索の基礎の上に作成されたもので、表面的には材料が豊富で多元的ですが、その実は羅列でしかなく、知的発展に与える影響はほとんどありません。日本には優れた文献学の伝統がありますから、こうした弊害が生まれる危険性は比較的低いはずです。私は実際にそうであることを望んでいますし、また台湾の若手研究者が日本の人文学研究のこうした側面から多くを学ぶよう望んでいます。私の本日の発表はここまでです。みなさまのご教示を賜りたいと思います。

訳注

[1] 中研院内部、およびそれに準ずる形での公開分を指す。

[2] 寒泉（http://140.122.127.253/dragon/）など。

[3] http://saturn.ihp.sinica.edu.tw/~wenwu/ww.htm

[4] http://idp.bl.uk/

[5] http://www.gale.com/EighteenthCentury/

訳者付記

本稿は、2004年3月20日に花園大学で開催された漢字文献情報処理研究会2004年春期公開講座の講演原稿に一部修正を加えたものである。翻訳にあたっては読みやすくするために見出しを付け、適宜改行を入れた。

付録：中文全文データベース目録（完成分）

2003年12月までの完成分、約299,470,000字

（作者付記：この書目分類は若干誤りを含んでいる可能性があります。）

◇壱．史書：131,019,861字

一、廿五史：39,969,529字

二、明清実録：62,480,565字

明実録
明実録附録
明実録校勘記
清実録（現在は資料庫の外部に設置）

三、その他：28,569,767字

古本竹書紀年輯証
国語
戦国策
前漢紀
後漢紀校注
八家後漢書輯注
東観漢記
九家旧晋書輯本
呉越春秋
越絶書
逸周書
列女伝
資治通鑑
続資治通鑑長編
建炎以来繫年要録
契丹国志
大金国志校証
華陽国志校補図注
水経注疏
水経注校釈
荊楚歳時記
東京夢華録注
入唐求法巡礼行記
読通鑑論
宋論
熹廟諒陰記事
元代白話碑集録
世本八種
宋稗類鈔
三朝北盟会編
宋元方志叢刊
馬可孛羅行記
張徳輝嶺北紀行足本校注
真臘風土記校注
元朝名臣事略
蒙古秘史新訳並註釈
歳時習俗資料彙編
四庫全書総目提要

◇弐．経書と子書：共21,798,677字

一、経書：9,335,206字

㈠ 十三経注疏：7,804,027字

㈡ 断句十三経経文：791,489字

㈢ その他：739,690字

点校四書章句集注
大戴礼記
韓詩外伝
読四書大全説

二、諸子と専著：10,932,292字

㈠ 諸子：4,104,777字

老子
老子校釈
荀子
晏子春秋
晏子春秋集釈
管子
管子軽重篇新詮
荘子
荘子集釈
荘子集解
荘子集解内篇補正
墨子
墨子間詁
墨子城守各篇簡注
韓非子
孫子
呉子
尉繚子
鬼谷子
尹文子
慎子
燕丹子
公孫竜子
鄧析子
関尹子
鶡冠子
通玄真経（文子）
孔叢子
淮南子
列子集釈
抱朴子内篇校釈
抱朴子外篇

㈡ 専著：6,827,515字

商君書
六韜
司馬法
呂氏春秋
孔子家語
春秋繁露
新序
説苑
新書
新語
新語校注
塩鉄論
法言義疏
方言校箋
論衡校釈
潜夫論箋校正
申鑒
中論
風俗通義校注
九章算経点校
周髀算経
釈名
顔氏家訓集解
斉民要術校釈
朱子語類
四存編
日知録
能改斎漫録
夢渓筆談校証
容斎随筆
敬斎古今黈

◇参．宗教文献：共12,035,363字

一、仏教：11,723,870字

㈠ 大正蔵：11,400,387字

第一巻　阿含部上
13　長阿含十報法経

14　仏説人本欲生経
23　大楼炭経
31　仏説一切流摂守因経
32　仏説四諦経
36　仏説本相猗致経
46　仏説阿那律八念経
48　仏説是法非法経
57　仏説漏分布経
91　仏説婆羅門子命終愛念不離経
92　仏説十支居士八城人経
98　仏説普法義経
第二巻　阿含部下
105　五陰譬喩経
109　仏説転法輪経
112　仏説八正道経
114　仏説馬有三相経
115　仏説馬有八態譬人経
131　仏説婆羅門避死経
137　舎利弗摩訶目連遊四衢経
140　阿那邠邸化七子経
149　仏説阿難同学経
150A　仏説七処三観経
150B　仏説九横経
151　仏説阿含正行経
第三巻　本縁部上
152　六度集経
○153　菩薩本縁経
154　生経
157　悲華経
167　仏説太子慕魄経
○170　仏説徳光太子経
184　修行本起経
○185　仏説太子瑞応本起経
186　仏説普曜経
190　仏本行集経
第四巻　本縁部下
196　中本起経
197　仏説興起行経
198　仏説義足経
200　撰集百縁経
○201　大荘厳論経
202　賢愚経
204　雑譬喩経
208　衆経撰雑譬喩
209　百喩経
210　法句経
211　法句譬喩経
212　出曜経
第八巻　般若部四
221　放光般若経
○222　光讃経
223　摩訶般若波羅蜜経
224　道行般若経
225　大明度経
第九巻　法華部全華厳部上
○263　正法華経
○266　仏説阿惟越致遮経
第十巻　華厳部下
280　仏説兜沙経
○285　漸備一切智徳経
第十一巻　宝積部上
313　阿閦仏国経
第十二巻　宝積部下涅槃部全
322　法鏡経
348　仏説大乗方等要慧経
350　仏説遺日摩尼宝経
356　仏説宝積三昧文殊師利菩薩問法身経
361　仏説無量清浄平等覚経
○362　仏説阿弥陀三耶三仏薩楼仏檀過度人道経
第十三巻　大集部全
398　大哀経
417　仏説般舟三昧経
418　般舟三昧経
第十四巻　経集部一
458　文殊師利問菩薩署経
492　仏説阿難問事仏吉凶経
506　犍陀国王経
525　仏説長者子懊悩三処経
526　仏説長者子制経
551　仏説摩鄧女経
553　仏説㮈女祇域因縁経
554　仏説柰女耆婆経
第十五巻　経集部二
602　仏説大安般守意経
603　陰持入経
604　仏説禅行三十七品経
605　禅行法想経
607　道地経
608　小道地経
621　仏説仏印三昧経
622　仏説自誓三昧経
624　仏説伅真陀羅所問如来三昧経
626　仏説阿闍世王経
630　仏説成具光明定意経
○632　仏説慧印三昧経
○638　仏説超日明三昧経
第十六巻　経集部三
684　仏説父母恩難報経
701　仏説温室洗浴衆僧経
第十七巻　経集部四
724　仏説罪業応報教化地獄経
729　仏説分別善悪所起経
730　仏説処処経
731　仏説十八泥犁経
732　仏説罵意経
733　仏説堅意経
735　仏説四願経
778　仏説菩薩内習六波羅蜜経
779　仏説八大人覚経
784　四十二章経
○790　仏説孛経抄
791　仏説出家縁経
792　仏説法受塵経
807　仏説内蔵百宝経
第二十二巻　律部一
1428　四分律
第二十三巻　律部二
1435　十誦律
第二十四巻　律部三
1467　仏説犯戒罪報軽重経
1470　大比丘三千威儀
1492　仏説舎利弗悔過経
第二十五巻　釈経論部上
1508　阿含口解十二因縁経
第二十八巻　毘曇部三
1557　阿毘曇五法行経
第四十七巻　諸宗部四
1985　鎮州臨済慧照禅師語録
1986A　筠州洞山悟本禅師語録
1986B　瑞州洞山良价禅師語録
1987A　撫州曹山元証禅師語録
1987B　撫州曹山本寂禅師語録
1988　雲門匡真禅師広録
1989　潭州潙山霊祐禅師語録
1990　袁州仰山慧寂禅師語録
1991　金陵清涼院文益禅師語録
1992　汾陽無徳禅師語録
1993　黄竜慧南禅師語録
附黄竜慧南禅師語録続補
1994A　楊岐方会和尚語録
1994B　楊岐方会和尚後録
1995　法演禅師語録
1996　明覚禅師語録
1997　円悟仏果禅師語録
1998A　大慧普覚禅師語録
1998B　大慧普覚禅師宗門武庫
1999　密菴和尚語録
2000　虚堂和尚語録

第四十八巻　諸宗部五
2001　宏智禅師広録
2002A　如浄和尚語録
2002B　天童山景徳寺如浄禅師続語録
2003　仏果圜悟禅師碧巌録
2004　松老人評唱天童覚和尚頌古従容庵録
2005　無門関
2006　人天眼目
2012A　黄檗山断際禅師伝心法要
2012B　黄檗断際禅師宛陵録
第四十九巻　史伝部一
2027　迦葉結経
2034　歴代三宝紀
2035　仏祖統紀
2036　仏祖歴代通載
第五十巻 史伝部二
○ 2042　阿育王伝
2059　高僧伝
2060　続高僧伝
2061　宋高僧伝
2063　比丘尼伝
第五十一巻　史伝部三
2066　大唐西域求法高僧伝
2076　景徳伝灯録
2087　大唐西域記
第五十二巻　史伝部四
2102　弘明集
2103　広弘明集
2104　集古今仏道論衡
2105　続集古今仏道論衡
2106　集神州三宝感通録
第五十三巻　事彙部上
2121　経律異相
2122　法苑珠林
第五十五巻　目録部全
2145　出三蔵記集

㈡　その他：323,483 字

洛陽伽藍記校注
六祖壇経
神会語録

二、道教：311,493 字

太平経合校

◆肆．医薬文献：30,046,879 字

金匱要略
難経本義新解
傷寒論
黄帝内経
本草綱目
衛生家宝産科備要
洗冤集録校訳
肘後備急方
外台秘要
医心方
太平聖恵方
普済方
医方類聚
名医別録
東垣医集
丹渓医集
赤水玄珠全集
遵生八箋
名医類案
続名医類案
脈経
黄帝針灸甲乙経
太平恵民和剤局方
子和医集
瘟疫論
神農本草経
千金翼方
備急千金要方
諸病源候論校注
婦人良方大全
宋以前医籍考

◆伍．文学と文集：55,933,113 字

一、総集：8,934,409 字

楚辞補注
敦煌変文集新書
文選
漢魏南北朝墓誌彙編
先秦漢魏晋南北朝詩
全上古三代秦漢三国六朝文
遼金元宮詞

二、別集：16,119,818 字

典論
祖堂集
亭林詩集
亭林文集
王梵志詩
顔元集
四庫輯本別集拾遺
全唐文
全宋詩
元代珍本文集彙刊—桐江集
全遼文
滋渓文稿
掲傒斯全集
湛然居士文集
元詩選
国朝文類
遼代石刻文編
元好問全集
鄭思肖集
元宮詞百章箋注
永和本薩天錫逸詩
謝畳山全集校注

三、歴代史料筆記叢刊：13,001,040 字

㈠《唐宋史料筆記叢刊》：3,593,325 字

隋唐嘉話
朝野僉載
明皇雑録
東観奏記
大唐新語
唐語林校証
東斎記事
春明退朝録
澠水燕談録
帰田録
竜川略志
竜川別志
東坡志林
黙記
燕翼詒謀録
涑水記聞
東軒筆録
青箱雑記
斉東野語
癸辛雑識
邵氏聞見録
邵氏聞見後録
桯史
遊宦紀聞
旧聞証誤
鉄囲山叢談
四朝聞見録
春渚紀聞
蘆浦筆記
鶴林玉露
湘山野録、続湘山野録
玉壷清話
泊宅編
老学庵筆記
西渓叢語
家世旧聞
石林燕語
雲麓漫鈔

鶏肋編
清波雑志校注
(二) 《元明史料筆記叢刊》：2,868,437字
南村輟耕録
草木子
菽園雑記
帰潜志
万暦野獲編
水東日記
戒庵老人漫筆
典故紀聞
玉堂叢語
寓圃雑記
穀山筆塵
四友斎叢説
治世余聞
継世紀聞
松窓夢語
広志繹
今言
三垣筆記
庚巳編
客座贅語
賢博編
奥剣編
原李耳載
玉鏡新譚
(三) 《清代史料筆記叢刊》：6,539,278字
池北偶談
蕉廊脞録
古夫于亭雑録
分甘余話
広東新語
郎潜紀聞初筆、二筆、三筆、四筆
郷言解頤
呉下諺連
巣林筆談
庸間斎筆記
異辞録
夷氛聞記
旧典備徴
安楽康平室随筆
広陽雑記
聴雨叢談
陶廬雑録
北游録
永憲録
揚州画舫録
蕉軒随録、続録
鏡湖自撰年譜
漏網喁魚集
海角続編
水窗春囈
世載堂雑憶
萇楚斎随筆、続筆、三筆、四筆、五筆
海国四説
不下帯編
巾箱説
清秘述聞三種
嘯亭雑録
枢垣記略
帰田瑣記
冷廬雑識
浪跡叢談、続談、三談
柳南随筆、続筆
履園叢談
竹葉亭雑記
簷曝雑記
四、小説戯曲：17,877,846字
山海経校注
世説新語箋疏
新校捜神記
穆天子伝
西京雑記
古小説鉤沈
児女英雄伝
西遊記
水滸全伝
繍像金瓶梅詞話
紅楼夢校注
儒林外史
三遂平妖伝
醒世姻縁
遊仙窟
老乞大諺解
朴通事諺解
訓世評話
閲微草堂筆記
桃花扇
型世言
警世通言
喩世明言
醒世恒言
三国演義
夷堅志
続夷堅志
揮麈録
子不語
続子不語
初刻拍案驚奇
二刻拍案驚奇
元刊雑劇三十種
歧路灯
永楽大典戯文
関漢卿戯曲集
新刊大宋宣和遺事
大唐三蔵取経詩話
鏡花縁
折獄亀鑑訳註
宋人軼事彙編
太平広記
◇陸．政書・類書と史料彙編：48,640,244字
漢官六種
通典
唐律疏議
唐令拾遺
唐会要
芸文類聚
文献通考
続文献通考
清朝文献通考
清朝続文献通考
太平御覧
通制条格
元代法律資料輯存
吏学指南
廟学典礼
秘書監志
五代史平話
宋大詔令集
建炎以来朝野雑記
宋朝事実類苑
大元聖政国朝典章
大清会典事例
元代奏議集録
站赤
憲台通記正続集
南台備要
*** 宋会要輯稿（四川大学古籍整理研究所、ハーバード大学東アジア言語文明学科と共同で完成。約10,277,585字）

2004年公開講座報告
デジタルアーカイブと東洋学

春期公開講座レポート

師　茂樹（もろ　しげき）

◆ はじめに

2004年3月20日、京都の花園大学で開催された漢字文献情報処理研究会2004年度春期公開講座「デジタルアーカイブと東洋学――真の活用を目指して――」について、簡単な報告を行いたい。

本公開講座においては、本会の小島浩之氏のコーディネートにより、台湾中央研究院で「漢籍電子文献」[1]の構築に携わってこられた陳弱水氏の講演が行われた。「漢籍電子文献」と言えば、本会でも『電脳中国学』などを通じてこれまで何度も紹介してきたし、本誌の読者であれば知らぬ者はいないであろう。

したがって、こちらから講演を依頼する際、「漢籍電子文献」の内容や利用方法などについて、言い換えれば技術的なことやHow-to的なことについて話していただくのではなく、このデータベースがどのような政治力学や学術的パースペクティブのもとに立ち上げられたのかについて話して欲しいと注文した。それが奏功したのか、講演はその期待に違わぬ「局内人（insider）」にしか語れない内容であった。

講演の内容については陳氏自身による本誌掲載の論稿を読んでいただくとして、本報告では聴衆としての感想や補足的なことを中心に述べていきたい。一部、重複する部分がある点はご海容いただきたい。

◆ 漢籍資料庫の発足当時と現況

講演では最初に、「漢籍電子文献」の発足当時（80年代中ごろ）のことが述べられた。その中で、人文学者と情報科学者の共同で始まったことが強調されていたのが非常に印象的であった。このようなコラボレーションは、当然なされてしかるべきことのようにも思われるが、実際、どれだけのプロジェクトがそれをなしえているだろうか。多くは人文学者が乏しい技術で悪戦苦闘しているか、情報科学者が人文学の方法論を知らないまま、技術的には高度かもしれないが実際にはあまり役に立ちそうにないものを作っているか、そのような場合が多いのではないだろうか。人文学者から見れば情報科学者は便利なコンピュータ屋さんでしかなく、情報科学者から見た人文学者は、自分の目的や方法論をきちんと説明できないわがままな人たち…と言うと言いすぎかも知れないが、それほどまでに異分野、ましてや文系と理系（という分類は好きではないが）の壁を越えたコミュニケーションというのは難しいものである。これが、コンピュータ利用の黎明期になされていたとは、驚愕せざるを得ない。

これに続けて、本誌2号に掲載された陳氏の論稿[2]以降の、「漢籍電子文献」の状況が紹介された。なんとこの間に1億4千万字から3億字弱へと倍増したとのことで、今後も増加するとのことである。その一例として個人的にうれしかったのは、『道蔵』の入力計画が「漢籍電子文献」において再スタートするとの報であった。

◆ 漢籍電子文献と中央研究院との関係

次のパートで述べられたのは、中央研究院という巨大組織における「漢籍電子文献」プロジェクトの位置づけ、およびプロジェクト運営における

苦労話であった。プロジェクトマネジメントについては、報告者もいろいろと苦い思いをしてきた経験があり、興味深く聞かせていただいた。

「漢籍電子文献」は元々、中央研究院の組織的な取り組みではなく、ましてや国家プロジェクトでもない、研究者の自発的な取り組みとしてスタートしたものだと言う。その後、「漢籍電子文献」が国際的に評価され成功したのをうけて、中央研究院本隊や政府からの研究費も出るようになったのだと言う。あれだけのデータベースが個人ベースで始まったと言うのも示唆的であるが、興味深かったのは、その成功を受けて、中央研究院の他の部門でもデータベース計画が行われるようになったものの、ほとんど成功していないという話である。「漢籍電子文献」は、中央研究院・史語所の研究者の自発性と自腹を切ってきたことの賜物であることが強調される反面、中央研究院全体におけるマネージメントの問題が指摘されており、どこの国でも同じだなと思った次第であるが、一方我が身を振り返って、「漢籍電子文献」の足元にも及ばない自分を恥じ入るばかりである。

また、中国で行われている『四庫全書』等のプロジェクトについては、「我々は資料そのものに興味があるのであって、何か特定の図書（をデジタル化すること）に興味があるのではない」という姿勢が述べられ、「漢籍電子文献」が伝統的な文献研究の延長線上にあることを強調していたのが印象的であった。これに関連して、質疑応答の時間に、中研院は古典の校訂にも著作権が存在すると認識しており[3]、テキストの選択や校訂の面で法的な問題の処理が様々あったことが紹介された。

報告者は、確率統計的な方法を用いた（日本では僻目で見られることもある）文献分析などにも取り組んでいる関係上、この陳氏の「伝統」の強調についてはひっかかるところもあった。しかし、後に氏の言う「伝統」が日本ではかなり革新的なことだということがわかった。すなわち、講演の最後で陳氏は、東洋学にとってデジタル化は必須であり、それ以外の選択肢はないと断言したのである。コンピュータ利用が「伝統」に正当に位置づけられる学風を、うらやましく思った次第である。

◇ 当事者としての回想

最後のパートでは、陳氏が３年にわたって責任者を務めた当事者としての回想が述べられた。

「漢籍電子文献」のアドバンテージとして、スタートが早かったこと、継続してやってきたこと、人文系の専門の研究者が携わってきたことなどをあげる反面、創生的な研究が評価される中央研究院内では、研究ツールである「漢籍電子文献」が評価されてこなかった点や、研究者がメンテナンスをしようとしない点に苦言を呈された。陳氏の言葉を借りれば、中央研究院の研究者はひとりひとりが「老板」だとのことである。

研究ツールという点に関して、台湾の東洋学界が工具書の類を日本・中国に負ってきた事を指摘し、「漢籍電子文献」が初めてツールの作成として貢献したことを強調していたのも興味深かった。この意識は、電脳化に遅れている日本の現状を考えると、そのうち逆転しそうな感じすらする。

◇ 最後に

以上、雑駁な報告であったが、本会がこれまで訴えてきたことにお墨付きをもらったような感じがする一方、単なるデータ量や技術ではない、学術上の高い目的意識を垣間見たことで、我が国に雨後の筍のように発生している表面的な電子化に対する危機感が一層募ってしまった。

注

[1] http://www.sinica.edu.tw/ftms-bin/ftmsw3

[2] 陳弱水（野村英登訳）「中央研究院歴史語言研究所漢籍全文自動化計画の発展、現状、未来」（『漢字文献情報処理研究』2、2001）

[3] 校訂の著作権については、今後、本会の著作権講座や本誌の連載などにおいて取り組む予定である。

東洋学情報化と法律問題——第2回

収蔵作品へのアクセスと法

石岡　克俊（いしおか　かつとし）

1. 序—著作権と所有権の峻別、そして所有権の射程

前回[1]、書籍・雑誌、レコード・CD、DVDやゲームソフトなど、著作物をさまざまな媒体に固定し内蔵している物、また、絵画、写真や彫刻など著作物と媒体とが一体化している物については、所有権者であっても、その物を全面的にかつ排他的に支配することができない場合があることを指摘した。

これは、著作権と所有権がそれぞれその客体を異にしているため、これらの物品において両権利の併存を許し、それがためにもたらされた状況だと理解することができる[2]。

このような考え方をはじめて明確に述べたのが、有名な「顔真卿自書建中告身帖」事件上告審判決[3]である。この事件において、所有権とはその所有物を使用・収益・処分する権利であるとされるところ（民法206条）、「美術の著作物」の原作品の所有者が、当該原作品の複製物の製作・販売を、その所有権を根拠として差止め・廃棄が可能か否かが争われた。

判決は、著作権と所有権を明確に峻別し、その所有権は「第三者が…原作品に対する排他的支配権能をおかすことなく原作品の著作物の面を利用した」ならば、侵害されたことにはならないことを示し、本件において問題となった原作品の複製物の製作・販売行為は、原作品の排他的支配権能を脅かすものではなく、所有権侵害を構成しないと判断した。

もとより、中国・唐代の著名な書家である顔真卿の筆跡は、著作権はすでに消滅したものと考えてよく[4]、かかる場合において、第三者がこの美術の著作物の複製物を出版すると、原作品の所有権者へ許諾を求める者が減少し、その結果、所有権者は収益をあげる機会が奪われ、経済上の不利益を受けることとなるかもしれない。しかし、この第三者による複製物の出版行為は、既に公有に帰した著作物を利用するに過ぎず、また、仮にこの行為が原作品所有者の経済上の不利益をもたらしたとしても、これは第三者が著作物を自由利用できる事実上の結果なのであるから、この行為を捉えて所有権者の原作品に対する使用収益機能を違法におかしたとは認められないと判示しているのである[5]。

この判決で示された留保条件（「第三者が原作品の排他的支配権能をおかすことなく」）の意味は、最近の判決である「かえでの木」事件[6]において詳細に言及される。すなわち、排他的支配権能の侵害の有無（損害賠償請求の根拠）につき、「かえでの木」事件判決は、「現に本件かえでの枝を折るなど本件かえでの所有権を侵害する行為を行なったとは認められない」[7]と認定していることからも明らかだが、専ら本件かえでの所有権侵害、しかも本件かえでに対する直接的・物理的侵害のことを指すものと考えられる。

他方、石岡[2003b]では、所有者の排他的支配権能をおかすこととされ、所有権侵害を構成すると判断された事例として、「広告ガス気球」事

件[8]と「長尾鶏」事件[9]とを指摘した。

「広告ガス気球」事件は、気球が広告媒体として使用することで利益を上げることを目的としているところ、第三者がこの目的に先立ち、気球を撮影した写真を素材とするポスターを別の広告宣伝に利用したために所期の目的が達成困難となった事例である。ここでは、排他的支配権能、すなわち使用収益機能の侵害の態様が、具体的目的との関係で詳細な認定を行なっている一方で、判決後半部分では、それと同じウェイトで予見可能性ないし過失論の中で損害賠償責任を排除できるロジックが用意されており、また、本件においては特有の事情がかなり考慮されたように見受けられ、本判決の射程はかなり限定的であるとの評価をした。

また、「長尾鶏」事件は、それ自体独特の美しさを備え、その管理・飼育にも工夫と苦労を要し育てられた長尾鶏にあって、その写真を絵葉書等に複製・販売した事例である。ここでは、管理・飼育に対する一定の努力を所有権との関係で法的に保護すべきものと捉えているようであるが、本件は、当該複製販売行為それ自体が不法行為を構成するか否か争われた例ではないことから、この先例的意義はかなり乏しいことを指摘した。

このように所有権の射程の問題は、その排他的支配機能の侵害と裏表の関係にあるが、これまで判例で明らかにされているのは、せいぜい「顔真卿自書建中告身帖」事件と「かえでの木」事件で示された点のみであり、他の二つの事件は真に所有権の問題といえるのか未だ不確定のところが多分にあることは否めない。近時においては、「クルーザー」事件[10]のように、所有権の内容たる排他的支配権について何ら言及することなく、「本件クルーザーの所有者として、同艇の写真等が第三者によって無断にその宣伝広告等に使用されることがない権利」という所有権とは別個の権利を観念し[11]、これに対する不法行為の成立を認め、理論構成する事案も出てきており、「広告ガス気球」事件や「長尾鶏」事件は、むしろこの延長線上で語られることが多い。

◇ 2. 収蔵作品管理の法的根拠

すでに取り上げた「顔真卿自書建中告身帖」事件において、最高裁判所は、博物館・美術館における料金徴収の根拠を「博物館や美術館において、著作権が現存しない著作物の原作品の鑑賞や写真撮影をするのに許可を要するとしているのは、原作品の有体物の面に対する所有権に縁由するもの」とし、「料金の徴収等の事実は、所有権が無体物の面を支配する権能までも含むものとする根拠とはなりえない。料金の徴収等の事実は、一見所有権者が無体物である著作物の複製等を許諾する権利を専有することを示しているかのようにみえるとしても、それは、所有権者が無体物である著作物を体現している有体物としての原作品を所有していることから生じる反射的効果にすぎない」と述べている[12]。つまり、美術館や博物館が収蔵品につき一定の条件を付したり、料金などを徴収するのは、収蔵品それ自体の所有権に基づくことが明らかにされているのである。

他方、「かえでの木」事件判決において「念のために」ではあるが、以下のようなかたちで、本件かえでに関する所有権以外の法的利益侵害可能性につき検討が行われている。

すなわち、(i)原告はある時期までは本件かえでを鑑賞するために平穏な態様で原告所有の土地へ立ち入ることを一般に容認していたものと認められること、(ii)本件かえでの保全を図るため、看板を立て、根を踏む等の本件かえでの生育に悪影響を及ぼす行為や、営利目的での土地への立ち入り及び撮影に制限を設けるなどの一定の条件を設定したが、観光客による本件かえでの鑑賞や、私的な目的での撮影のために土地に立ち入ることを何ら禁止していなかったこと、(iii)被告が当該看板設置以降撮影をしていないこと、また、本件かえでの生育に悪影響を及ぼす可能性のある行為をしていないことを認定し、「本件土地」の所有権侵害の可能性を問題にしている[13]。

「かえでの木」事件判決のように「本件土地」の所有権に目を向けることは、博物館・美術館等

における収蔵品へのアクセスや博物館・美術館等の料金徴収の法的根拠について一定の示唆を提供する。

つまり、この事件の判決を前提とすれば、収蔵品へのアクセスの法的根拠は、収蔵品それ自体の所有権に基づいているのではなく、実は、美術館や博物館の土地の所有権（管理権）に由来しているということが、容易に理解されるのである[14]。

これで、本論稿の主な論点である「収蔵作品へのアクセス」の問題は、収蔵品の所有権に基づくという理解に立つにせよ、収蔵品を所有する美術館・博物館等の土地の所有権（管理権）に由来するという理解に立つにせよ、著作権ではなく所有権という民事上の権利との関係で理解され、取扱われるべき問題であるということが明らかとなった。

しかし、「収蔵作品へのアクセス」に関する法的問題を考える場合、これまで指摘してきた論点の検討だけではいまだ十分とはいえない。民事上の問題として処理されるのであれば、収蔵品へのアクセスの問題は、私的自治原則のもと、主として当事者間の合意によって規律されることになるのは論を俟たないだろう。だが、わが国のみならず多くの国々では、文化政策的視点から、歴史的に貴重な文献や資料、また美術品・文化財の類は、国や地方自治体などの公的機関が管理している場合が圧倒的に多い。そうなると、私人間（しじんかん）において妥当する考え方とは異なる法的思考が要求され、公が関わることに伴うさまざまな配慮が必要となる。

そこで、以下では「収蔵作品へのアクセス」につき公的機関が関係した具体的な事例を紹介し、その主要な論点を順次取り上げ、解説を加えていくことにしたい。

3. 収蔵作品へのアクセスに関する判例―「県立美術館収蔵作品非公開等措置」事件

3.1. 事案の概要

本件は、Y_1（被告・富山県）が開設した富山県立近代美術館（以下､「県立美術館」という）が、X_1（原告）の製作した「遠近を抱えて」と題する連作版画[15]（以下､「本件作品」という）を購入・収蔵していたところ、本件作品及び本件作品を収録した図録（以下、「本件図録」という）を非公開、また本件図録を非売品とし（以下、「本件非公開措置」という）、さらに、本件作品を売却し（以下､「本件売却」という）、本件図録を焼却した。そこで、本件作品の作者である X_1 は表現の自由を侵害されたなどとして、また、X_1 を除くその他Xらは知る権利を侵害されたなどとして、Y_1（被告・富山県）に対し損害賠償（X_1 は 100 万円、その他Xらは各々 10 万円）を求め（原審について [3.2.3] 及び [3.2.4] 参照、控訴審について [3.3] 参照）、さらに、Y_2（被告・富山県教育委員会教育長）に対し本件売却及び本件焼却の無効確認（原審について [3.2.1] 参照）ならびに本件作品の買い戻し及び本件図録の再発行（義務づけ）（原審について [3.2.2] 参照）を求めた事案である。

なお、X_1 は、昭和 24 年に富山県で生まれた美術作家で、大学を卒業後、本件作品を含む連作版画「遠近を抱えて」14 点を始め、多数の版画等を製作し、国内外の展示会で作品の展示を続けてきた。また、その他Xらはいずれも、本件作品の鑑賞、本件図録の閲覧又は購入を希望していた。

他方、県立美術館は、昭和 56 年に開館し、国内外における 20 世紀美術の流れを展望するとともに、郷土美術の伝統をたしかめ発展させることを基本姿勢とし、21 世紀をめざして、新しい創造の可能性を見いだすにふさわしい文化拠点としての役割を果たすことを運営の基本方針としている。

県立美術館は、美術品等を収集、保管、展示することや、美術に関する案内書・解説書・目録・図録・年報・調査報告書等を作成し、頒布する事業を行い（富山県近代美術館条例4条⑴、⑷）、県立美術館においては、常設展示室及び企画展示室において展示している美術品を観覧することができるほか、Y_2の許可を受けて、展示又は保管している美術品について学術研究等のために模写・模造・撮影等の特別観覧をすることができ、Y_2は、美術品の管理のために必要な範囲内で条件を付すことができるとされている（同条例7条1、2項）。ただし、特別観覧をしようとする者は、特別観覧をしようとする日の7日前までに、特別観覧許可申請書をY_2に提出しなければならない（同条例施行規則4条1項）。

さらに、同条例は11条で入館の拒否及び制限について規定し、⑴他の入館者に迷惑となる行為をするおそれがあるとき、⑵施設設備、美術品又は美術資料を汚損し、又は損傷するおそれがあるときには、Y_2は入館を拒否できるとしている。

⦿ 3.2. 原審・判旨

以下、原審・地裁判決 [16] のポイントを判決にそって整理する。

● 3.2.1. 無効確認訴訟の適否について

Xらの訴えの一つは、Y_2に対する本件売却及び本件焼却の無効確認を求めるものであり、これは行政事件訴訟法3条4項にいう無効確認訴訟である。同法にいう「行政庁の処分その他公権力の行使に当たる行為」とは、法律行為ばかりではなく事実行為も含まれる。また、行政庁の法令に基づく行為のすべてを意味するものではなく、公権力の主体である国・地方公共団体が行う行為のうち、その行為によって直接、国民の権利義務を形成し又はその範囲を確定することが法律上認められているものをいう。

したがって、本件売却は富山県立近代美術館美術品管理要綱及び富山県会計規則に基づき、本件焼却は富山県会計規則に基づき、それぞれなされているところ、同要綱ないし同規則は、いずれも美術品等の取得ないし処分等につきY_1ないし県立美術館等の内部的手続を定めているにすぎず、他に、本件売却及び本件焼却について国民の権利義務を形成し又はその範囲を確定することを定めた法令は存在していない。したがって、「本件売却及び本件焼却は、直接Xらの権利義務を形成し又はその範囲を確定することが法律上認められている場合に該当するものということはできず、行政庁の処分には当たら」ない [17]。つまり、XらのY_2に対する本件売却及び本件焼却の無効確認の訴えは、不適法である [18]。

● 3.2.2. 義務付け訴訟の適否について

Xらは、Y_2に対し本件作品の買い戻し及び本件図録の再発行を求めているが、このような義務付け訴訟が許容されるのは、(i)行政庁の処分要件が一義的に明確であり、裁量の余地がないこと（明白性）、(ii)事前審査を認めないことによる損害が大きく、事前救済の必要性が顕著であること（緊急性）、(iii)他に適切な救済手段がないこと（補充性）といった要件を満たす場合に限られるとされている。本件において、県立美術館による美術品の購入及び図書の発行に関しては、美術品等の取得ないし処分等に関する手続について富山県会計規則及び富山県立近代美術館美術品管理要綱に規定があるほかは、どのような美術品を購入し、どのような内容の図録を発行するかにつき、規定は存在せず、これらはその性質上、美術に関する高度な専門的判断が必要であり、Y_2の広範な裁量に委ねられている。このことから、本件義務付け訴訟は、少なくとも明白性の要件を欠き、この訴えは不適法である [19]。

● 3.2.3. 本件非公開措置等の違法性の有無について

本件非公開措置等が違法性の有無は、詰まるところ、Xらが主張するXらの知る権利等が本件非公開措置等により侵害されたか否かの判断に帰することになると考えられるが、X_1主張の「本件作品を市民に鑑賞してもらう権利」、「本件図録に掲載された本件作品を市民に鑑賞してもらう権

利」のいずれも認めることが出来ず、本件非公開措置等により X_1 の名誉感情が著しく傷付けられたこと、その作品発表の場が大きく制約されたことも認められない。

また、その他Xら主張の「本件作品を鑑賞し、図録を閲覧する権利」については、表現の自由（憲法21条1項）ないし知る権利の保障の枠内で考えられるが、知る権利は、国民が公権力に対して情報の開示を求めるという積極的権利の面においては、それが公権力に対して作為を求めるものである場合、法令による開示基準の設定と具体的開示請求権の根拠付けがあってはじめて、裁判規範性を有する。

富山県立近代美術館条例にいう特別観覧制度は、その意味で、県立美術館に収蔵されている作品についての知る権利を具体化する趣旨のものである。この条例において不許可の場合についての規定はないものの、県立美術館は地方自治法244条1項にいう公の施設に当たり、正当な理由がない限り、特別観覧許可申請を不許可としてはならず、拒んだ場合には、憲法が保障する知る権利を不当に制限することになると解するべきであるとされる（図録の閲覧については、同条例において閲覧を認める規定はないものの、県立美術館が所蔵する図録を閲覧することは公の施設を利用することに当たり、特別観覧制度と同様、正当な理由がない限り、住民が図録を閲覧することを拒んではならず、拒んだ場合には憲法で保障する知る権利を不当に制限することになる。)）。

したがって、「本件作品の特別観覧許可申請を不許可とし、本件図録の閲覧を拒否しうるのは、本件作品を特別観覧させ、あるいは、本件図録を閲覧させることによって、他者の基本的人権が侵害され、公共の福祉が損なわれる危険がある場合に限られるものというべきであり、このような場合には、その危険を回避し、防止するために、本件作品の特別観覧ないし本件図録の閲覧が必要かつ合理的な範囲内で制限を受けることがあるといわなければならない」[20]。

その際、本件の非公開措置等が、原告Xらの有する知る権利に対する必要かつ合理的な範囲内の制限であるか否かについては、次の２つの視点から検討が行われる。

一つは、本件作品が昭和天皇の写真をコラージュしたものであったことから、本件作品が昭和天皇のプライバシーの権利ないし肖像権を侵害するか否かという観点から、いま一つは、本件作品及び本件図録を公開することに伴う公開派及び非公開派双方からの抗議行動のためもたらされる美術館職員の業務の支障ないし美術館としての平穏の確保という観点から、である。

前者につき、天皇の象徴としての地位及び職務から、天皇にはプライバシーの権利や肖像権の保障は制約されると考えられ、それを踏まえると、本件作品が昭和天皇のプライバシーや肖像権を侵害し又はその疑いがあるとは認められない。

また、後者につき、Yの主張するように美術館が管理運営上の障害を理由として作品及び図録を非公開とすることができるのは、利用者の知る権利を保障する重要性よりも、美術館で作品及び図録が公開されることによって、人の生命、身体又は財産が侵害され、公共の安全が損なわれる危険を回避、防止することの必要性が優越する場合である。本件につき客観的事実に照らして見ると、本件作品及び本件図録の公開により、これらの利益が損なわれ、明らかな差し迫った危険の発生が具体的に予見されたものということはできない。

したがって、本件作品の特別観覧許可申請の不許可、本件図録の閲覧の拒否は違法なものというべきである。

● 3.2.4. Y_1 の責任

「本件作品の特別観覧許可申請の不許可は、Y_2 が県の職務を執行するについて行なった違法な侵害であり、本件図録の閲覧拒否は県立美術館館長が県の職務を執行するについて行なった違法な侵害であり、少なくとも過失があったものといわざるを得ないから、Y_1 は国家賠償法１条により右侵害行為によって生じた損害を賠償する責任がある」[21]。

＊　＊　＊

以上、一審判決は、Y_2に対する無効確認訴訟、義務付け訴訟をいずれも却下し、Y_1に対する損害賠償請求を一部認容した。つまり、県立美術館が収蔵品を非公開とする等の措置をとった場合において、観覧希望者が作品の特別観覧許可申請をして不許可とされ、また作品を収録した図録の閲覧を求めて拒否されたときに、知る権利の違法な侵害にあたるとし、県（Y_1）に損害賠償責任を認めたのである。

これを受け、請求を棄却されたXら及びY_1が控訴したのが、つぎで取り扱う控訴審判決[22]である。

⊙ 3.3. 控訴審・判旨

本判決は、一審判決と同様の理由で、Y_2に対する無効確認訴訟及び義務付け訴訟は不適法であるとした。また、本件非公開措置等の違法性については、X_1との関係では、一審判決とほぼ同様の理由で、違法性は認められないとし、X_1以外のXらとの関係では、以下のように判示し、原判決中、Y_1敗訴部分を取消し、Y_1に対する請求を棄却し、Xらの控訴を棄却した。

ここでは、後の検討の便宜上、一審判決と異なった判断が示された部分を取り上げ、整理しておくことにしよう。

● 3.3.1. 県立美術館条例と「知る権利」

「県立美術館についての…条例等の規定は、美術館の開設趣旨やその規定の仕方、内容に照らしても、…憲法21条が保障する表現の自由あるいはそれを担保するための「知る権利」を具体化する趣旨の規定とまで解することは困難である。しかしながら、県立美術館は地方自治法244条1項にいう公の施設に当たり、県立美術館が所蔵する美術品を住民が特別観覧することは、公の施設を利用することにほかならないから、Y_2は、地方自治法244条2項に定める正当な理由がない限り、住民のした特別観覧許可申請を不許可とすることは許されないと解すべきである」（なお、本件図録の閲覧についても同様の判示をしている）。

● 3.3.2. 「正当な理由」（地方自治法244条2項）

「Y_2による本件作品の特別観覧許可申請の不許可、県立美術館及びY_2による本件図録の閲覧の拒否について、地方自治法244条2項の「正当な理由」が認められるか否かについて検討するに、県立美術館としては、購入・収蔵している美術品や自らが作成した美術品の図録については、…観覧あるいは閲覧を希望する者にできるだけ公開して住民への便宜（サービス）を図るように努めなければならないことは当然であるが、同時に美術館という施設の特質からして、利用者が美術作品を鑑賞するにふさわしい平穏で静寂な館内環境を提供・保持することや、美術作品自体を良好な状態に保持すること（破損・汚損の防止を含む。）もその管理者に対して強く要請されるところである。これらの観点からすると、県立美術館の管理運営上の支障を生じる蓋然性が客観的に認められる場合には、管理者において、右の美術品の特別観覧許可申請を不許可とし、あるいは図録の閲覧を拒否しても、公の施設の利用の制限についての地方自治法244条2項の「正当な理由」があるものとして許される（違法性はない）というべきである」。

● 3.3.3. 県立美術館に対する管理運営上の支障の蓋然性

「本件作品及び本件図録の本件非公開措置が県立美術館の管理運営上の支障を生じる蓋然性が客観的に認められる場合に該当するか否かについて検討する。…本件作品及び本件図録の公開について、県立美術館等に対し、執ような抗議、抗議文の送付、県立美術館館長等との面談の要求、本件作品等の廃棄や県立美術館長の辞任等を求める右翼団体による街宣活動、富山県立図書館における本件図録の破棄事件、県知事に対する暴行未遂事件などが相次いで発生しており、これらの状況のもとで本件作品及び本件図録を公開（図録の販売を含む。）をした場合には、県立美術館の管理者としては本件作品の特別観覧者及び本件図録の閲覧者を含めた利用者に平穏で静寂な環境を提供・保持する要請を満たすことができなくなる可能性

が多分にあり、また、特別観覧制度を利用して本件作品を損傷しようとする者が紛れ込む可能性が否定できない状況にあったというほかはないから、県立美術館の管理運営上の支障を生じる蓋然性が客観的に認められる場合に該当するものと認めるのが相当である。そうすると、その余の点について検討するまでもなく、本件非公開措置には、地方自治法 244 条 2 項に定める「正当な理由」が有るというべきであるから、違法性は認められず、…Y_1 に対する損害賠償請求は理由がない」。…以上のとおりであるから、X らの本訴各請求中、Y_1 に対する請求はいずれも理由がないから棄却すべきであり、Y_2 に対する請求に係る訴えはいずれも不適法として却下すべきである。

◇ 4. 論点の整理と解説

⊙ 4.1. 美術品等の取得・処分の法的性格

● 4.1.1. 行政庁の処分性

X らの、Y_2 に対する無効確認の請求の中で、公立美術館・同博物館の収蔵作品売却ないし収蔵作品掲載図録焼却の法的性格がある程度明らかになった。

本件における収蔵作品の売却は県立美術館美術品管理要綱及び県会計規則に基づき、また、本件における収蔵作品を掲載した図録の焼却は県会計規則に基づきそれぞれなされ、これらはいずれも美術品等の取得・処分等についての内部的手続を定めている。

無効確認訴訟（行政事件訴訟法 3 条 4 項）の対象となる「行政庁の処分その他公権力の行使に当たる行為」とは、判決にも明らかにされているとおり、法律行為に加え事実行為も含まれ、行政庁の法令に基づく行為のすべてを意味するものではないものの、国・地方公共団体などの公的機関が行う行為のうち、その行為によって直接、国民の権利義務を形成し又はその範囲を確定することが法律上認められているものをいう。

本件において問題となった美術館等の収蔵作品の売却及び収蔵作品掲載図録の焼却について、国民の権利義務を形成し又はその範囲を確定することを定めた法令は存在しておらず、当該行為は、直接国民の権利義務を形成し又はその範囲を確定することが法律上認められている場合に該当しないから、行政事件訴訟法上の「行政庁の処分」には当たらないものとされた。したがって、これらの行為について、行政庁に一定の責任を追及するためには、収蔵作品の売却等行為につき何らかの実体法的根拠が必要だということになる。しかし、次に見るように、美術館における美術品の購入等については、高度に芸術的専門的判断の必要性があることからその権限の行使にあたっては或る程度の裁量が予定され、詳細な根拠付けは、一般論として期待できないであろう。

● 4.1.2. 高度な専門的判断に基づく広い裁量

また、X らが求めた作品の買い戻し及び図録の再発行を求めるいわゆる義務付け訴訟の可能性が提起されたが、その中で、県立美術館による美術品の購入及び図書の発行に関する高度な芸術的専門的判断の必要性が指摘され、それゆえに広範な裁量を有するものと評価された。

一般に、公的機関に対し一定の作為を義務付ける訴訟（いわゆる義務付け訴訟）においては、(i)明白性、(ii)緊急性、(iii)補充性といった要件を満たすことが必要とされている。

本件においては、県立美術館による美術品の購入及び図書の発行に関しては、美術品等の取得ないし処分等に関する手続について県会計規則及び県立美術館美術品管理要項に規定があるほかは、どのような美術品を購入し、どのような内容の図録を発行するかにつき、規定は存在せず、これらはその性質上、美術に関する高度な専門的判断が必要であり、Y_2 の広範な裁量に委ねられているとする。このことから、本件義務付け訴訟は、すでに少なくとも明白性の要件を欠くことになるので、認められないこととされた。

通説によれば、義務付け訴訟が認められるのはかなり例外的であることが前提とされていることから、既に指摘した要件のもと、本件事情を踏まえた判断は是認されよう。義務付け訴訟の適否の

評価はさて措き、ここで重要なのは、美術館による美術品の購入等につき、高度な芸術的・専門的見地からの判断が性質上必要とされ、そのために広範な裁量が認められることが明らかにされている点である。

◉ 4.2. 本件条例及び県立美術館の法的位置づけ

原審では、国民の権利ないし自由を公権力との関係で規律する日本国憲法との関連について言及されている。すなわち、その他Xらの主張にある「本件作品を鑑賞し、図録を閲覧する権利」については、表現の自由（憲法21条1項）ないし知る権利の保障の枠内で考えられ、知る権利は、国民が公権力に対して情報の開示を求めるという積極的権利の面においては、それが公権力に対して作為を求めるものである場合には、法令による開示基準の設定と具体的開示請求権の根拠付けが伴えば、裁判規範として援用することができるとする。その上で、特別観覧制度を定める県立美術館条例に触れ、同条例が、県立美術館の収蔵作品についての知る権利を具体化する趣旨のものであると位置づけている。

しかし、控訴審判決では、県立美術館についての条例等の規定は、その内容に照らしても、憲法21条が保障する表現の自由あるいはそれを担保するための「知る権利」を具体化する趣旨の規定とまで解することは困難であるとし、憲法上保障される表現の自由ないし知る権利との関係を否定する。

本件条例をどのように捉え位置づけるかという問題である。これには、原審のように憲法上の権利（知る権利）と関連づけて考える場合と、一応、それとは分離して理解する控訴審の考え方とが可能だが、この捉え方の違いが、次に述べる本件非公開措置等（収蔵品へのアクセス拒絶）の当否の判断において異なった対応を導くことになる。

他方、かかる違いが認められながらも、県立美術館が地方自治法244条1項にいう公の施設[23]に当たることは双方とも認めており、県立美術館の所蔵美術品の特別観覧は公の施設を利用することにほかならず、Y_2は「正当な理由がない限り」（地方自治法244条2項）、住民の特別観覧許可申請を不許可とすることは許されないと解している（なお、本件図録の閲覧についても同様）。

◉ 4.3. 本件非公開措置等（収蔵品へのアクセス拒絶）の当否の判断

原審判決は、この条例には不許可の場合についての規定がないが、県立美術館は、正当な理由がない限り、特別観覧許可申請を不許可としてはならず、拒んだ場合には、憲法が保障する知る権利を不当に制限することになると解するべきである（図録の閲覧についても同様とする）。したがって、本件作品の特別観覧許可申請を不許可とし、本件図録の閲覧を拒否しうるのは、これらの行為により、他者の基本的人権が侵害され、公共の福祉が損なわれる危険がある場合に限られ、このような場合には、その危険を回避し、防止するために、本件作品の特別観覧ないし本件図録の閲覧が必要かつ合理的な範囲内で制限を受けることがあることを認める。

このことから、本件の非公開措置等が、原告Xらの有する知る権利への必要かつ合理的な範囲内の制限であるか否かについて、一般的には、次にような判断基準を提示している。

すなわち、(i)利用者の知る権利を保障する重要性よりも、美術館で作品及び図録が公開されることによって、人の生命、身体又は財産が侵害され、公共の安全が損なわれる危険を回避、防止することの必要性が優越する場合であり、かつ、(ii)その危険性の程度は、単に危険な事態を生ずる蓋然性があるというだけでは足りず、客観的な事実に照らして、明らかな差し迫った危険の発生が具体的に予見されることが必要である、と。これは、公共施設における集会等の不許可事由として最高裁判所により示された判断基準である。憲法21条によれば、集会の自由は表現の自由の中に含まれ、精神的自由と位置付けられる。これらの権利を制限するには、より慎重な配慮が求められることから、これらの要件が判例上形成された。このように、原審判決は、収蔵品へのアクセス拒絶を表現

の自由との関連で捉えている。

他方、控訴審判決によれば、地方自治法244条２項の「正当な理由」の解釈につき、美術館という施設の特質を考慮し、(i)利用者が美術作品を鑑賞するにふさわしい平穏で静寂な館内環境を提供・保持することや、(ii)美術作品自体を良好な状態に保持すること（破損・汚損の防止を含む。）も管理者に対して強く要請され、これらの観点から、県立美術館の管理運営上の支障を生じる蓋然性が客観的に認められる場合には、美術品の特別観覧許可申請を不許可とし、あるいは図録の閲覧を拒否しても、公の施設の利用の制限について「正当な理由」があると述べられている。つまり、管理者の側に、施設の特性上要請される直接的かつ合理的な利益と、施設管理運営上の必要性が客観的に認められる範囲内で、「正当な理由」ありとされ、原審判決で示されたような憲法上保護される権利との比較衡量はここでは排除されている（そればかりか、原審で示された判断基準は、「憲法21条が保障する『集会の自由』を制約するおそれのある事案については相当であるが、本件のような美術品及びその図録の観覧あるいは閲覧に関する事案については厳格に過ぎ、相当ではない」とし、最高裁判例を本件に適用した地裁判決の判断を否定している。）。

5. まとめ

前回の考察に続き、著作権と所有権の峻別が主な論点となった「顔真卿自書建中告身帖」事件や、所有権の射程について重要な示唆を提供するいくつかの判決を簡単にレビューし、美術館や博物館等における収蔵作品の法的取り扱いについて整理・検討を加えた。

また、美術館や博物館等が、多くの場合、公的機関として位置づけられることに鑑み、かかる特性から生ずる法的諸問題を、近時、話題となった「県立美術館収蔵作品非公開等措置」事件の一審判決及び控訴審判決を紹介した。ここでは、一定の行為が公権力によってなされた場合、表現の自由や知る権利を保障する憲法、あるいは行政事件訴訟法や国家賠償法に代表される行政法が、どう関わり合いをもってくるのかを判例を通してかいま見てきた。

取り上げた判例は、法学上（行政法上）、数多くの興味深い論点を有している。しかし、当然のことではあるが、ここではそれを網羅的に検討することはしてはいない。「収蔵作品へのアクセスと法」というタイトルの本稿の関心は、もっぱら公的機関が有する収蔵品の法的位置づけ、かかる機関による美術品等の取得・売却の法的性質、そして収蔵品へのアクセス拒絶に関する正当化根拠にあるからである。

なお、本稿で紹介した考え方やこれらの議論の延長線上には、いわゆる美術品のみならず、歴史的に貴重な文献やさまざまな発掘資料など公的機関が収蔵しているものの公開を求める場合にも同様の考え方が妥当するということを、付け加えておくことにしたい。

参考文献

1. 塩野宏 [2001]『行政法Ⅱ』（有斐閣、第２版、2001年）
2. 石岡克俊 [2001]『著作物流通と独占禁止法』（慶應義塾大学出版会、2001年）
3. 石岡克俊 [2003a]「著作権法に基づく権利の行使と競争秩序―頒布権・消尽・独占禁止法」法学研究（慶應義塾大学）76巻１号（2003年）
4. 石岡克俊 [2003b]「所有権の行使と無体財産の法的保護―判例の分析と解説」漢字文献情報処理研究（漢字文献情報処理研究会編、好文出版刊）４号（2003年）
5. 石岡克俊 [2004]「消尽理論の法と経済学」林紘一郎編著『著作権の法と経済学』所収（勁草書房、2004年）

注

[1]　石岡 [2003b]4 頁以下。

[2]　より厳密には、所有権の客体は有体物であるのに対し、

著作権のそれは無形の思想・感情の表現といった観念的な無体物であり、有体物の物理的な支配は可能であっても、その有体物に固定され内蔵されている観念的な無体物までをも支配することは不可能であるということによる。なお、このような違いから、著作権法には、絵画、写真及び彫刻などの「美術の著作物」につき、著作権と所有権との間の調整を企図した規定が存在する。たとえば、展示権（著作権法25条）に関し、「美術の著作物」をその原作品の所有者等がその所有権に基づき公に展示することを認めている（屋外の場所に恒常的に設置する場合を除く。）規定（著作権法47条）がそうである。また、知的財産権法上の法理である消尽理論も両権利間の調整を企図したものとしての理解が可能である（詳細は、石岡 [2003a] 及び石岡 [2004] 参照。）。

[3] 書籍所有権侵害禁止請求事件、最高裁昭和58年（オ）171号、昭和59年1月20日第2小法廷判決、上告棄却(民集38巻1号1頁､判例時報1107号127頁)。

[4] 近代的な著作権という概念それ自体がこの時代においてはそもそも存在していなかったはずである。

[5] 「顔真卿自書建中告身帖」事件・書籍所有権侵害禁止請求事件、最高裁昭和58年（オ）171号、昭和59年1月20日第2小法廷判決（判例時報1107号127頁〔128-129頁〕）。なお、後に触れる「かえでの木」事件においても、この判決の考え方が基本的に踏襲され、本件かえでに対する所有権の内容として、それは、もっぱら有体物としての排他的支配権能にとどまり、本件かえでの写真の複製や複製物を掲載した書籍の出版についての排他的支配権能を含まないことが明らかにされ、併せて、第三者による本件かえでの写真の複製や複製物を掲載した書籍の出版・販売によって本件かえでの排他的支配権能が侵害されない限り、本件かえでの所有権は侵害されないことも示されている。

[6] 「かえでの木」事件・出版差止等請求事件、東京地裁平成14年（ワ）1157号、平成14年7月3日民事29部判決、棄却（確定）（判例時報1793号128頁。）

[7] 「かえでの木」事件・（判例時報1793号128頁〔132頁〕。）

[8] 控訴審：損害賠償等請求控訴事件、東京高裁昭和52年（ネ）792号、昭和53年9月28日民事8部判決（著作権判例集ⅲ 846頁、肖像権関係判例集859頁）、原審：損害賠償等請求事件、東京地裁昭和48年（ワ）7540号、昭和52年3月17日民事15部判決、棄却（控訴）（判例時報863号64頁）。

[9] 損害賠償請求事件、高知地裁昭和57年（ワ）31号、昭和59年10月29日判決、棄却（控訴）（判例タイムス559号291頁）。

[10] 損害賠償等請求事件､神戸地裁平成元年（ワ）265号、平成3年11月28日判決、一部認容一部棄却（確定）（判例時報1412号136頁）。

[11] これらは、近年、いわゆるモノの肖像権として整理されている。

[12] 「顔真卿自書建中告身帖」事件・書籍所有権侵害禁止請求事件、最高裁昭和58年（オ）171号、昭和59年1月20日第2小法廷判決（判例時報1107号127頁〔129頁〕）。

[13] 「かえでの木」事件・（判例時報1793号128頁〔132頁〕。

[14] なお､「かえでの木」事件判決は「本件土地」の所有権侵害可能性を認める実益として次のように指摘する。すなわち､「原告は、本件かえでの所有権に基づき…各行為を阻止できない限り、本件かえでを保全することができない旨述べる。しかし、原告が、本件土地上に所在する本件かえでの生育環境の悪化を憂慮して、本件かえでの生育等に悪影響を及ぼすような第三者の行為を阻止するためであれば、本件土地の所有権の作用により、本件かえでを保全する目的を達成することができる。…現に、原告は、本件土地への立ち入りに際しては、本件かえでの生育等に悪影響を及ぼす可能性のある行為をしてはならないこと、許可なく本件かえでを営利目的で撮影してはならないことを公示しているのであるから、第三者が上記の趣旨に反して本件土地へ立ち入る場合には、原告は当該立入り行為を排除することもできるし、上記第三者には不法行為も成立する。また、本件土地内に、美観を損ねないような柵を設けること等によって、より確実に上記目的を達成することもできるというべきである」(「かえでの木」事件･(判例時報1793号128頁〔133頁〕。）と。

[15] この作品は、昭和天皇の写真と裸婦像などを組み合わせたコラージュ連作版画であった。

[16] 国家賠償等請求事件､富山地裁平成6年（ワ）242号、平成10年12月16日民事部判決（判例時報1699号

120 頁)。

[17] 同上〔135-136 頁〕。

[18] 同上〔136 頁〕。

[19] 同上〔136 頁〕。

[20] 同上〔138 頁〕。

[21] 同上〔140 頁〕。

[22]「県立美術館収蔵作品非公開等措置」事件控訴審判決・国家賠償等請求控訴事件、名古屋高裁金沢支部平成 11 年（ネ）17 号、平成 12 年 2 月 16 日第一部判決、取消（上告）(判例時報 1726 号 111 頁)。

[23] 公の施設とは、地方自治法の定めるところによれば、地方公共団体が住民の福祉を増進する目的をもってその利用に供するための施設であるとされる（塩野宏 [2001]171 頁参照)。

漢字文献情報処理研究会　第 7 回大会のお知らせ

本会第 7 回大会は、文部科学省科学研究費特定領域研究「東アジア出版文化の研究」G 班との共同企画として、下記要領で開催することになりました。

文部科学省科学研究費特定領域研究「東アジア出版文化の研究」G 班セミナー
漢字文献情報処理研究会第 7 回大会　ジョイント企画

セミナー「漢籍の情報化—これからの出版文化—」

日時：2004 年 12 月 4 日 ㈯ 13:00-17:00
場所：関西大学尚文館 1 階マルチメディア AV 大教室

- 12:00 ～ 12:30　漢字文献情報処理研究会総会（漢情研会員のみ）
- 13:00 ～ 14:30　研究発表

「漢籍コーパスの歩みと現状」山田崇仁（独立行政法人日本学術振興会特別研究員 PD）
「漢籍テキスト処理の現状と展望」秋山陽一郎（京都大学人文科学研究所 COE 技術補佐員）
「漢籍と電子出版」野村英登（東洋大学講師）

- 15:00 ～ 17:00

パネルディスカッション　「漢籍の新しい形—ネットワークとアーカイブ—」

⦿司会

二階堂善弘（関西大学助教授）

⦿パネラー

相田　満（国文学研究資料館助手）
石岡克俊（慶応大学産業研究所助教授）
守岡知彦（京都大学人文科学研究所附属漢字情報研究センター助手）
師　茂樹（花園大学専任講師）

- 17:30 ～ 19:30　懇親会（関西大学生協 BONPLAT）

2004年夏期公開講座報告
東洋学情報化と著作権問題 II

法理論と実務の狭間

「東洋学情報化と著作権問題 II」から

小島　浩之（こじま　ひろゆき）

◇ はじめに

2004年度の漢情研夏期公開講座は、昨年に引き続いて[1]「東洋学情報化と著作権問題 II」と題し、7月19日（月）に早稲田大学国際会議場において開催した。

講師は石岡克俊氏（慶應義塾大学産業研究所助教授）と笹井真也氏（丸善学術情報ナビゲーション事業部企画担当部長）のお二人にお願いした。石岡氏には、昨年度に引き続いて講師をご快諾いただき、また笹井氏には本業のお忙しい最中に、無理を言って業界内の様々な話題を提供していただいた。まずは主催者および参加者を代表して御礼申し上げる。

なお筆者は今年度、公開講座担当幹事として春期、夏期の講座ともに裏方に回っていた。従って本稿が純粋な参加レポートではなく、講座の企画段階からの問題等も含めた回顧のような形になることをお許し願いたい。

◇ 守成の業は難しい？？

本会として著作権を題材とした講座は二度目であり、どのような形で行うか、大変難しい問題であった。昨年は、参加者がはじめて経験する著作権講座という前提で、石岡氏には、法学の基本知識や著作権法の基礎的な考え方の部分に、かなりの時間を割いていただいた。また講師への質問をBBSで募集したところ多数寄せられ、当日の質問に対する解説、議論も大いに盛り上がった。

漢情研執行部としては今年度も二匹目のドジョウを狙おう？と考えたのだが、これが簡単ではなかった。昨年の参加者も今年度初めての参加者も、どちらも満足できるような講座構成をという課題にぶつかったからである。「創業は易く守成は難し」と言うように、この講座一つをとっても継続して確乎たるものにしてゆくことの難しさを感じた。

思案した後、講演を石岡・笹井両氏の二本立てとし、内容に幅をもたせることにした。石岡氏は昨年度の参加者から継続して話を聞きたいとの要望が多かった。昨年度の参加者の為には石岡氏の続編を盛り込むべきだろう。他方、笹井氏のように参加者全員が初めて話を聞く方にも講演していただければ、以前の参加の有無に関わらず、参加しやすいはずである。また石岡氏による著作権の法構造からの分析に加え、実務家である笹井氏に実務処理上の問題点を明らかしていただくことは、非常に興味深い内容・議論になるだろう。こうして担当者の不安な気持ちは一転して大きな期待に変わった。

ところが講座構成の確定後、別の問題が発生した。開催日が三連休の最終日となり、参加しにくい日程になった。また昨年同様BBSで質問を募集したが、ほとんど反応が無く担当者を慌てさせた。これらは一つに担当者である筆者の準備不足もあり、反省している。全ての対応が後手後手に

回ってしまい、講師のお二人にはご迷惑をお掛けした。心よりお詫び申し上げる。

最終的には昨年には及ばないものの、20 名以上の参加者があり、予想した以上に密度の濃い議論になり、執行部一同胸をなで下ろしている。

以上前置きが長くなったが、さっそく当日の講演のレポートに移ることにしよう。

◇「著作権と所有権の相互関係から所蔵権（？）へ」

最初の 80 分で標記の講演題目で石岡氏に講演いただいた。石岡氏の講演は本誌前号に寄稿された論考 [2]（以下、石岡論文）を元に、さらに議論を進めるという形式であった。

⦿ 判例の読み方

石岡論文には、多くの判例が引用されている。筆者のような法律の素人には、判例の文章は難解であり、石岡氏の解説の字面を追うのがやっとであった。しかし今回の講座が、判例における記述の約束事 [3] に触れながら、石岡氏と一緒に読むという形で進められたおかげで、おぼろげに判例の文意が理解できたような気がした。歴史史料の読解が、専門家の手ほどきを必要とするように、判例の読解も同様だという印象を受けた。

⦿ 著作権と所有権の調整

この二つの権利については、混同されることが多い。古い本では法学者ですら区別できていないというから、素人が間違えるのは仕方のないことかもしれない。

著作権と所有権の対象を区分した有名な判例が“「顔真卿自書建中告身帖」事件” [4] であり、客体の相違に基づく両権利の併存を示したものとして特筆されるという。両者の相違を簡単にまとめると次の表のようになる。

しかし、著作物を何らかの媒体に固定しているもものや、両者の一体化しているものは両権利の調整が必要となってくる。特に問題となるのは絵画・写真などの美術の著作物についてである。

	著作権	所有権
法　規	著作権法	民　法
規律対象	無体物 [5]	有体物 [6]
権利内容	発想そのものへの利用権	所有物を使用、収益及び処分する権利

石岡氏は二種類の調整の存在を指摘する。一つは“実定法”上の調整規定、これは文言として形をとっている（条文として法に書かれている）もの、もう一つは“判例法理”に基づく調整で、裁判所の判断を積み重ねたものだという。特に興味をひいたのは後者についての解説であった。ここでは「消尽理論」[7] を中心として、映画における頒布権や海外における追及権 [8] など著作権と所有権の間を調整する学説や諸権利について話していただいた。

著作権に関わる際には、対象物の有する諸権利をきちんと場合分けして考え、実定法だけでなく可能な限り法理の解釈にまで踏み込まなければ、最終的な判断はできない。おそらくこれは、著作権法に限らずどのような法律を考える場合でも同様なのであろう。

⦿ 民法の重要性

著作権の範囲だと思われがちなものでも、実は別の権利や別の法的根拠に基づくものが多くある。

- モノの写真等の掲載許諾は所有権の問題である。ただし判例によれば、写真の撮影は有体物に損害を与えるわけではなく、美術館などの「写真撮影禁止」は所有権を根拠とすることはできない。
- 判例の範囲で所有権を捉えれば、博物館や美術館における入場料徴収の根拠は、（所有権に基づく）モノに対するアクセスではなく、敷地に踏みいること（土地所有権）に拠っている。
- 電子媒体にみられるソフトの配布制限の規定は著作権ではなく、民法における契約の

問題である。

これらから言えるのは、著作権を考える場合、民法を無視できないということではないだろうか。講演の最後に富山県立美術館が非公開とした作品について、公開を求めた裁判を例にして、知る権利と所有権の問題について言及された（本号の石岡氏の論文で論じられるであろうから詳細は割愛する）。この話の中で興味を覚えたのは、憲法とは国と国民の関係を規定したもので、私的な問題には弱いということであった（知る権利じたいが憲法に基づくものであり、上記の裁判の場合は公的機関であるため、知る権利に基づき裁判となった）。裏を返せば、個人間の関係に属する問題については、民法こそが第一の法規なのだと言える。そもそも著作権法も、民法の規定する無体物に関しての特別法（こういった言葉が正しいかどうか解らないが）に値するものではないか。所有権、契約など様々な民法上の問題を理解して、はじめて著作権問題と向き合えるのかもしれない。

◆ 著作権処理の現場から

つづいて、笹井氏に60分間講演していただいた（特に演題が無かったため、標題は筆者が付した）。丸善は、様々な機関から資料のデジタル化業務を請け負っており、その際の著作権処理を一手に引き受けてこられたのが笹井氏である。本誌レビューでも採り上げられている、国立国会図書館の「近代デジタルライブラリー」[9]の著作権処理も氏の仕事だったという。

今回は2002年に刊行した「国立国会図書館所蔵　昭和前期刊行図書デジタル版集成　社会科学部門」（37,314件、46,670冊）の著作権調査と著作権処理の経験に基づき、お話いただいた。

◎ 誰のどういう権利を処理するのか？

著作権処理を行う際には、著作者を特定する作業から始めなければならない。今回は、図書館資料のデジタル化にともなう作業であったから、当然、カードなどの目録データがまず参照されることになる。

ところが図書館の作成する目録規則に従った書誌データでは、著作権者を明確にできないという。書誌データ上の著者だけが著作権者でない場合や、実際の著作権者と目録上の著者が異なる場合があるのだという[10]。これは、目録学・書誌学の立場から図書館学に関わっている筆者にとって、まったくもって「コロンブスの卵」であった。

確かに笹井氏の指摘のように、『日本目録規則』では、著者を全て記録しなくても良い場合（例えば著作者が複数にのぼる場合、代表著作者だけを記録して省略するなど[11]）が認められている。また、所定の情報源（タイトルページ、奥付、背、表紙）に記載されているかどうかで、著作者となったりそうでなかったりする場合がある[12]。このため目録規則に従うと採録されない著作権者が出てしまう。さらに挿し絵、表紙デザインなどの著作権者は図書じたいに記載のないことも多い。著作権者の表示が無くとも、著作権が存在するものは沢山あるのだ。目録規則はこれらに対応しているとは言い難く、著作権処理ということに限っては、図書館の目録データを鵜呑みにはできない。

そもそも出自の異なる目録規則と著作権法の体系を同じ俎上で論じることじたい難があるのかもしれない。しかし、著作物を専門的に扱う図書館における著者の概念が、法的な概念と異なるのはいらぬ誤解を生じる恐れもあるだろう。今後、著作権法上の著作権者の概念に基づき、目録規則上の責任表示の概念に手を加えることが必要になってくるかもしれない。

ところで、著作権者が確定しても、どの権利を処理すべきか問題になることがある。笹井氏の経験では、デジタル化したものをイントラネットで公開することは、複製権、公衆送信権のいずれに対して許諾を得るのか問題になったという[13]。当時の文化庁では担当課長の交代の度に解釈が変わり、大変苦労したとのことであった。

◎ 文化庁長官裁定制度の舞台裏

文化庁との折衝の苦労といえば、これが最たるものだろう。文化庁長官裁定制度とは、相当の努

力を払っても著作権者が確定できない場合、文化庁長官に裁定を仰ぎ、補償金を支払うことで著作権処理できるというものである。このように説明すると簡単に思えるが、“相当の努力”＝“並大抵の努力では駄目”だという。著作権者を捜しているという事実を、五大新聞や一般月刊誌で公表するなど、一個人では不可能な努力を要するという[14]。そもそも裁定制度は、規定があまりに曖昧であったり、担当者によって対応が異なったりと、ルールはあって無いようなものらしい。

笹井氏は､問題点に関して､文化庁との間で「交渉記録をとり、法解釈の根拠を問い記録し、これらを文書で再確認する」という三つのステップを踏むことで、事例を積み上げ、無法地帯？に一つのルールを作り上げたという。この作業で丸善が受けた裁定数は 8,938 点 9,133 冊[15]。笹井氏の地道な努力の積み重ねが功を奏したというべきだろう。

しかし笹井氏の積み上げたルールは、その多くがあくまで丸善と文化庁の間で解釈された結果である。そのまま行政としての公式見解となっていない部分や、今後の法改正や判例によって覆る部分もあるだろう。この状況では、一般の裁定制度利用はとうてい不可能だ。笹井氏は裁定制度について、最低限のルールが必要だと力説されていたが、全くもってその通りだろう。

◇ むすびにかえて

講座に参加された方はもちろんのこと、この拙稿をお読みなった方も、講座名と内容のギャップを感じておられるかもしれない。今年度の講座は、「東洋学情報化と著作権問題 II」と題しながら、東洋学に特化しているわけではないのである。しかし、昨年度の講座で不明瞭であった、裁定制度や著作権と所有権の関係、著作隣接権などの話題に特化したものとなっており、決して昨年度の講座との間に断絶があるわけではない。著作権問題をはじめとする知的所有権問題はもはや一部地域の問題ではなく、グローバルなものとなっている。したがって東洋学情報化における著作権問題も、著作権問題全体の流れの中で位置づけられなければならない。その意味では、本年度の講座は非常に有意義であったのではないだろうか。最後になったが当日ご参加いただいた多数の参加者に御礼申し上げる。

「継続は力なり」 来年度第３段が開催できるように願いつつ筆を措くことにしよう。

◇ 補足

昨年度ペンディングとなっていた著作権とその周辺権利に関わる問題の一つに、校訂権の問題があった。これについて講師に触れていただくことができなかった。このため後ほど、筆者が笹井氏にメールでお聞きした。ここに笹井氏の了解を得て、丸善の著作権処理における校訂権の取り扱いについて要点をメモしておく。

- 校訂者の問題は、厳密には「創作に相当する行為」（つまり「相当な加筆」の有無）があったかどうかに関わる。しかし編集内部の（ゲラ刷り校訂など）実際工程が分からないと判断できないことが多い。従って図書本体中から解る校訂者について検討した。
- 講義を筆記し、諸本により校訂したと思われるものは､校訂者を著作権者とみなした。
- 名前貸しと分かるものは、著作権者として扱わなくても良いが、この場合、文化庁に名前貸しである具体的な証拠を示さねばならなかった。
- 現在の出版社では、内部編集者の校訂が大量にあるので、これらを含め版面権（出版権）[16] と呼んでいるようだ。ただし現行著作権法では、著作者、学術研究校訂者、内部編集者という関係がある著作物であれば､前２者のみが著作権関連者。ただし「△△出版社編集部編」と明記されている場合は、出版社を著作権者として見なす。

注

[1] 昨年度（2003年度）の公開講座「東洋学情報化と著作権問題」については師茂樹「『東洋学情報化と著作権問題』参加レポート」（『漢字文献情報処理研究』4, 2003.10）を参照。なおこのレポートの簡略版が、「漢字文献情報処理研究会メールマガジン」第39号（2003.8.15）にも掲載されている。

[2] 石岡克俊「所有権の行使と無体財産の法的保護：判例の分析と解説」（『漢字文献情報処理研究』4, 2003.10）

[3] 例えば、「判例において原告および控訴人はX、被告および被控訴人はYと記述するため、第1審と第2審でXYが入れ替わる可能性がある、つまり1審の被告が控訴した場合、第2審の判例では1審の被告Yが2審の控訴人としてXと表記される。このことを理解していないと、一つの事件の判例を下級審から上級審へ順次見ていく場合に、事実関係を誤認する恐れがある。」ということや、「判決文の中に本論に対して傍論と呼ばれる部分のある場合があり、この部分は法的効力を持たない箇所である。」といったことなどをご教示いただいた。

[4] 判決の詳細は前掲注[2]石岡論文を参照。
顔真卿は唐代の人物で書家として有名、建中は唐の年号、告身は辞令書のこと。この作品は建中元年（780年）、顔真卿が太子少師に任命された時の告身で、彼の自書とされている。書のテキストとして著名であるが、素性の明らかな唐代の辞令書の一例として、唐代官僚制度の研究においても重要な史料である。

[5] 権利、自然力（電気・熱・光）、知的創造物、技術的思想などを指す。

[6] 物理的に一定の空間を占めるものを指す。

[7] 権利者が適法に生産していったん流通においた限り、その物を購入して使用、販売、貸与等する行為は権利侵害を構成しないとする法理。

[8] フランス・イタリア・アメリカカリフォルニア州（再販売追及権法）などが認めている権利。美術の著作権について、美術品の移転のたびごとに、著作権者が対価の一部を徴収できる権利。絵画は、描かれた後かなりの時間を要して市場価値の上がることが多いので、著作権者が不利にならないようこの権利が認められているという。
日本では美術市場がオープンでないため権利行使がしにくい。オークション制度が整わない国では制定しても意味がないという。

[9] 拙稿「図書館とOPAC」（『漢字文献情報処理研究』2, 2001.10）

[10] 戦前の図書には、出版法に基づいて出版社の社長名が著作権者として記載されていることがあり、実際の著作権者ではない場合があるという。

[11]『日本目録規則　1987年版　改訂二版』（以下NCR1987R）2.1.5.1D“一つの責任表示に記録する個人名や団体名の数が2までのときはそのまま記録し，3以上のときは，主なもしくは最初の名称一つを記録し，他は「[ほか]」と補記して省略する”
※目録規則では著者などのことを責任表示と表現する。

[12] NCR1987R 2.1.5.1の後半部分に“監修者, 校閲者, スポンサーとしての団体名等が所定の情報源に表示されているときは, これを責任表示の範囲に含める。”とあって、監修者や校閲者は、記載場所により著作者として認められる場合とそうでない場合があることが解る。また記載場所によっては、著作権法上は著作権者となり得ないスポンサーなどが著作者として扱われるということもあることになる。

[13] 現在では、インターネットでの公開やLANなどのイントラネット利用は複製権ではなく、送信可能化権として処理される

[14] 丸善の場合は、商業行為の一貫として許諾を求めたので、非営利の場合や個人の場合は多少対応が異なることも予想される。

[15] 丸善ではこの冊数の図書について3回に分けて裁定を申請した。文化庁側の数え方では、件数とは申請回数のことを指すようで、これだけ多くの処理が行われたにも関わらず公表件数としては3になるというから驚きである。

[16] 版面権（出版権）は、著作権法上認められていない。にも関わらず新聞社、出版社などが主張しているのは、認められることを期待しているふしがあるという。また最近では印刷社ですら主張する場合があるという。

特集 1

Wiki・Weblog と人文学

昨年来、インターネット上では Wiki や Weblog が大きな関心を呼び、大手ポータルサイトやプロバイダの中には利用者増の手段として Weblog サービスを提供する所もある。

Wiki とは WWW 上で動的にページの構築・更新が行えるサイト管理システムのことであり、複数人が共同で Web サイトを構築可能であるという特徴を持つ。

また Weblog（略して blog ともよぶ）は Web 日記をベースにしたシステムだが、「コメント」「トラックバック（一種の他の Weblog からのリンク）」といった双方向性を持つ点に特徴がある。

これらは従来の Web サイト・掲示板・メーリングリストといったコミュニケーションツールに対し、よりインターネットの「動的」「ネットワーク」「共同作業」といった特徴を生かしたシステムであるといえる。

本特集は、Wiki や Weblog を人文学や授業でどう活用するかについての試行錯誤の一端である。

読者諸賢におかれては、本特集を御一読いただき、ぜひとも Wiki や Weblog を活用していただきたい。

Contents

プロジェクト研究における Wiki の活用

田邉　鉄（たなべ　てつ）

◆ はじめに

本稿は、北海道大学大学院国際広報メディア研究科（以下本研究科）における実践を例に、共同作業環境としての WikiWikiWeb（以下 Wiki）[1] の有効性を検証する。

本研究科は高度職業人養成をうたう、実学系大学院である。そのためインターンシップや、企業・団体との連携講座など、社会連携に主眼を置いた教育プログラムを実施している。プロジェクトメソッドはその中核をなす教育方法論である。

プロジェクトメソッドは、専門分野の異なる複数の学生からなるプロジェクトチームで、社会の緊急な課題を取り上げ、問題を調査し、調査に基づいた具体的な提言を作成、あるいは提言の実践までを授業の枠内で行う。これまでに「雪印事件一年後のブランドイメージ」、「ニセコ町観光振興」などのテーマを取り上げ、「実効性のある提案」[2] との評価を受けている。これらのプロジェクト研究は、特別演習として授業化されているが、それ以外に、通常の授業、ゼミでもプロジェクトメソッドを採用することが奨励されている。

プロジェクトメソッドにおける教師の役割は、企業等外部のパートナーと最初のつなぎをつけることや、研究が迷走しそうになった時に、適宜助言を与えることなど、補助的なものにとどまる。筆者は主としてプロジェクトを円滑にすすめるための IT 環境を整備・運用するサポートを行った。その中で、一部の学生やアドバイザを務める教員から、使いやすいグループウェアを用意してもらえないだろうか、という相談があった。ほとんどのプロジェクトチームは、研究を進めるにつれて次第にメンバーが増えていく。後からの参加者にも、それまでの資料や議論をきちんと見てもらうためのライブラリ機能や、進捗状況を確認できるようなスケジュール機能、さらに対面での

図 1　Wiki の画面

編集(管理者用) | 編集 | 差分 | 新規作成 | 一覧 | RSS | FrontPage | 検索 | 更新履歴

FrontPage - YukiWikiへようこそ

目次

- ごあいさつ
- お願い
- 練習用Wikiサイト
- 使い方

YukiWikiへようこそ

ごあいさつ

このページはYukiWiki2のFrontPageです。

- RecentChangesは最近更新されたページの一覧です。
- YukiWikiのスクリプトのダウンロードは http://www.hyuki.com/yukiwiki/ からどうぞ。
- IndexPageは全ページの一覧です。
- YukiWiki2作業記録
- YukiWiki2不具合報告
- YukiWiki2.1新機能

議論や打ち合わせになかなか参加できない社会人学生のためのフォーラム機能などの要望があった。

筆者は最終的に、「Wiki を使うのが最も適当」という結論を伝えた。複数のプロジェクトでは Wiki が非常に活発に使われたが、より多くのプロジェクトで設けたもののほとんど使われなかったり、そもそも「設置自体必要ない」と言ってよこしたりした。

ここではまず、筆者が Wiki を使うのが適当と判断した理由を Wiki の特徴を紹介しながら述べる。その後で、失敗（または門前払い）のケースと成功のケースを取り上げ、Wiki がプロジェクト研究で効果を上げるための条件を探る。

◇ 1　Wiki とは何か

Wiki とはウェブブラウザさえあれば、誰でも、どこからでも編集したり、ページを追加したりできるウェブサイトである。ウェブアプリケーションとしては極めて単純なシステムでありながら、開発者をひきつけてやまない魅力があるようだ。Ward Cunningham が最初の Wiki を公開して以来、様々な機能を付け加えたクローンが多数登場している。ここではまず、Wiki の基本的な機能を紹介し、Wiki の魅力を探る。

図 1 はごく一般的な Wiki の画面である。普通のウェブページとなんら変わるところはないが、画面上部に「編集」というメニューがある。これをクリックすることで、ページの内容をその場で書き換えることができる（図 2）。誰もが参加できるという点では電子掲示板に似ているが、ページ全体を書き換えられることや、他人の書いた内容も含めて加除修正できるというところで異彩を放っている。

ページはタイトルとテキストで記述された内容が一対一で対応した一種のテキストデータベースとして保存される。閲覧する時には、このテキストデータから自動的に HTML が作成されるので、内容を更新するのに HTML の文法や FTP の使い方を覚える必要はない [3]。ウェブ上の掲示板を利用する知識さえあれば利用することができる。特に不特定多数、あるいは複数の団体からのメンバーが参加するようなプロジェクトでは、このエントリーバリアの低さは魅力的だ。

また、Wiki には簡単に使えるリンク機構がある。Wiki への書き込みの中に、一定のルールで埋め込まれたテキスト [4] は、ページタイトルと解釈され、自動的にそのタイトルを持つページへのリンクが生成される。

秀逸なのは、そのタイトルを持つページが存在しなくても、リンクが生成されることである。この場合、リンクをクリックすると、新しいページの編集画面があらわれる。これは「未決リンク」と呼ばれる仕組みで、「このキータームに関する情報を誰か知りませんか？」という問いかけといえる。未決リンクのおかげで、サイトを訪れる誰もが「このサイトには何が欠けているか」を即座に把握することができる。

Wiki のこの機能は、複数のメンバーで網羅的に情報収集するための画期的な方法と言える。取り込むべき情報を、Wiki のリンク形式で書き込

図 2　編集画面

んだページをひとつ用意しておけば、それは即、進捗状況をひと目で確認できる To-Do リストとなる。

どの Wiki にも備わっている機能はその程度だが、さて、読者のみなさんは Wiki についてどのような感想をお持ちだろうか。正直に言って、私は最初 Wiki にはあまりそそられなかった。ウェブブラウザからサイトを更新するツールや、たくさんの機能を盛り込んだグループウェアの類はいくらでもある。それらに比べて Wiki は明らかに見劣りする。

日本製 Wiki の草分け YukiWiki の開発者である結城浩氏は、この点について「Wiki はネットワーク上に置かれたホワイトボード」[5] なのだ、と説明する。かっちりしたコンテンツを公開するというよりは、雑多なメモ書き（時には落書き）を残し、あとから役に立つかどうか考える。いらなくなったら消してしまう。いかにも間に合わせのツールに見えるが、パソコンの周りにメモの紙片をちらかしてしまいがちな人なら、そういう間に合わせツールの有効性はすぐに理解できるのではないだろうか。

ちなみに、Wiki という変わった名前は、ハワイの言葉で「すばやい」という意味である。サクサク使えるツールという意味だろうが、Wiki にはもう一つ「ラフな」「間に合わせの」といった意味があるそうだ。

◆ 2 なぜ Wiki なのか

本研究科のプロジェクトメソッドをサポートする IT 環境に必要な条件はおおむね以下の通りである。

- 最終的に完成させるのは、Word による提案書、PowerPoint によるプレゼン資料、Excel による調査報告の 3 つである
- 研究の概要と成果は Web で公表する
- 複数のプロジェクトが同時並行で実施されるので、メンバー管理、スケジュール管理ができること
- 利用者教育に時間を割けないので、エントリーバリアが低いこと
- 外部パートナーや社会人学生の便宜をはかるため、どこからでも利用できること

これらの条件を満たすシステムとして、私の頭にあったのは 4 つのツールだった。電子メール、掲示板、市販のグループウェア、Wiki である。以下それぞれの検討ポイントを示し、Wiki 採用に至った経緯を述べる。

⊙ 電子メール

エントリーバリアが最も低く、外部パートナーとの連携にも好都合なのは、電子メールである。プロジェクト単位でメーリングリストを設けることにより、プロジェクトの進捗状況を確認したり、打合せの議事録などの情報を共有することが容易にできる。ただ、共同で手を入れている Word ファイル等をメールでやり取りするのは避けた方がよい。メールの添付ファイルは、ファイルのコピーを送っているに過ぎず、きちんとバージョン管理をしなければ、細かな違いを含むクローンが氾濫することになる。共同作業ツールとしては不完全で、単なる連絡ツールにしかならない。

⊙ 電子掲示板

本研究科は発足した 2000 年から、WebBBS という電子掲示板システムを導入し、授業の連絡や討論、イベントなど情報の交換、事務からのお知らせなど、広く活用されている。教員、学生全員が参加しているので、ここに、プロジェクトごとのスレッドを用意することで、簡易なグループウェアとして使えるのではないか、と考えた。

WebBBS はごく普通の電子掲示板だが、多数のスレッドやメンバーを管理する機能が充実している。また添付ファイルが使えるので、Word 等の文書ファイルを修正のたびにアップしておくだけで、版の管理ができる。共同研究のプラットフォームとして十分な機能を備えているといえる。ただ、クローズドなシステムであるため、外部パートナーを参加させる際に、特定の会議室だけのア

クセス権限を与えるなど細かな調整を手作業で行わなければならない。また、せっかく Web 上で情報を蓄積できるのに、一般公開用のページは別に作る必要がある。

⊙ 市販のグループウェア

市販のグループウェアのほか、それと同等の機能を持つ Web サービス [6]、同等の機能を実現できるフリーの Contents Management System なども含む。スケジュール管理や To-Do など、プロジェクト管理に使える機能や、ファイル交換、特定の事柄に関する討論を行うフォーラムなど、プロジェクト研究を進めるために必要な機能は網羅されている。

さらに、Microsoft Office SharePoint Portal Server のように、Word や Excel のファイルをネットワーク上で共有し、直接編集することが可能なものであれば、特に報告書やプレゼン資料を作成することが容易になる。メールや掲示板の添付ファイルとの大きな違いは、あるメンバーが現在手を入れているファイルは、必ず最新版である、という保証が得られることである。

これら製品版のグループウェアの弱点は、機能が高くなればなるほど、覚えなければならないルールが増えることである。説明書の記述ひとつとっても、ある製品では「参加者」となっているところが、別の製品では「プレイヤー」「メンバー」だったりする。そんな些細な違いでも、多数・多様なメンバーを含むプロジェクト内で統一するのは大変な作業になるだろう。

⊙ だから、Wiki

Wiki のエントリーバリアは高くないと言えるが、電子メールや掲示板に比べると、多少は特別なルールを覚える必要がある。また、メンバー管理機能はそもそもない [7]。これらの弱点を差し引いたとしても、「ホワイトボードに書かれた文章をみんなで見ながら叩く」というモデルをネットワーク上で実現できる、という大きなメリットには替えがたい。

学校で顔をあわせる学生たちにとって、ホワイトボードを前に打ち合わせをする、というのはリアルな話である。であれば、Wiki の作業スタイルにもすぐ馴染むと考えたのだが、現実はそうは甘くなかった。

◇ 3　失敗と成功はどこで分かれたか

最初に Wiki は使えない、と言い出したのは「ニセコ町の観光振興策」プロジェクトのリーダー格の学生である。彼女が言うには、「こういう新しいものを使うには、強力なリーダーシップでトップダウン式に進めるしかないが、20 名全員の合議制で方針を決めるようなチームではとても使いこなせない」そうだ。また、「まだ内容がまとまっていないのだから、載せるべきコンテンツがない」という声もあった。そういう段階だからこそ Wiki なのだ、と言ってはみたが、いらないといっているものを無理強いできるわけでもない。結局 Wiki を立ち上げたものの、半年たらずで閉鎖された。

この話から得られる教訓は、Wiki を使った仕事のやり方を直感的に理解できないと、誰も Wiki を使おうとは思わない、ということである。Wiki はシステムがもつ機能の多さではなく、全員に書き換え権限を与えるという思想や仕事のスタイルのユニークさによって評価されるシステムである。多機能なグループウェアのような「目に見える」便利さではないので、理解するのは難しい。Wiki の利用に「強力なリーダーシップ」が必要であるとは思わないが、少なくとも、Wiki に慣れた誰かが、いくつかのページを作って、実際の利用方法を示す必要があるだろう。

Wiki を使いたいといっていたのに、次第に使われなくなってしまったケースもある。「留学生に向けたウェブ広報」プロジェクトは、最終的な成果物が大学広報のためのウェブサイトであり、プロジェクト終了後も引き続きメンテナンスが必要であることから、最初から Wiki を使うことが検討されていた。大学の公式ウェブサイトから取り出した情報を、メンバーが分担して翻訳、

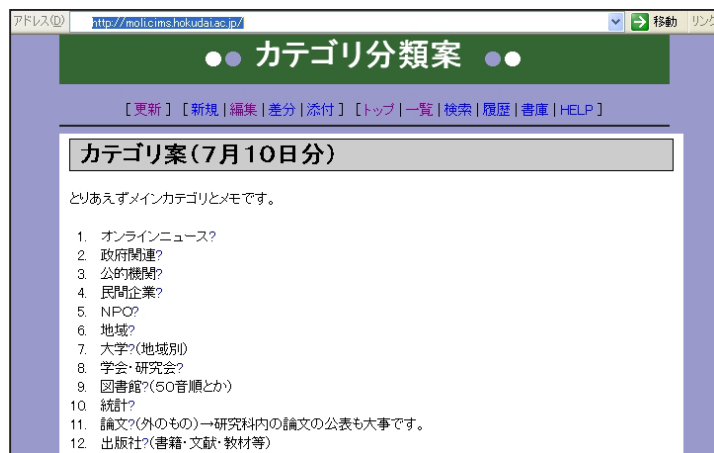



図 3　リンク集構築作業用 Wiki

Wiki に掲載して叩き、語彙や表現の統一などに配慮しながら、最後に公開するという手順で進められた。

この時は、最後の「公開」にクレームがついた。もともと、大学の学生プロジェクト推進経費を受けて進めていたプロジェクトなので、大学の公式ウェブサイトの内容を用いることに問題はない。ところが、Wiki のような動的なサイトではなく、静的なウェブサイトで、しかもなるべく公式サイトのページ構成やデザインを踏襲すべし、という条件がついた。大学の公式サイトは、学内委員会の責任で運営している。ボランタリーなサイトとはいえ、大学がオーソライズしたページに間違いがあったり、（公開時に編集機能を制限したとしても）故意に誤った情報に書き換えられたりしたら、責任問題になる、という、まあ国立大学（当時）としてはもっともな懸念ではある [8]。また、モニタ利用をお願いしていた台湾人学生からも、どこに何の情報があるのかわかりにくい、という意見が出た。広報・広告サイトを Wiki で置き換えるのは無理と判断し、単に翻訳作業のテキスト置き場として Wiki を使うことにした。

ところが、このプロジェクトはもともと人数が 3 名しかおらず、公開・メンテナンスまでをすべてを Wiki でやれないのであれば、導入の意味もない、と言い出し、結局 Wiki はほとんど使われないまま放置された。

これまでのプロジェクトの中で、Wiki の利用がもっとも効果を上げたのが、「研究科リンク集構築プロジェクト」である。リンク集は静的なウェブページとして作るか、またはフリーウェアとして配布されているスクリプトを流用することが決まっていたので、Wiki はプロトタイピングと、集めた URL の分類、整理だけに用いられた。まず、研究科の学生、教員等からウェブブラウザの「お気に入り」（Book Mark）を提供してもらい、Wiki に貼り付けた。それと並行して、カテゴリ別、場所別、授業科目別などの分類を、Wiki のリンク機構を利用して作り、そこに集まった URL を貼り付けていった。

図 3 はカテゴリ分類に関する議論を行っている最中のページであるが、そこで作ったカテゴリ分類案からそのまま各カテゴリへのリンクが貼られ、未決リンクとなっている。あとはリンクを開いて、実際に URL を貼り付けてみるだけで、プロトタイピングができる。このように、種類の違う情報を関連付けて仕事を効率よく進めることができたのは、Wiki ならでは、と言えるだろう。

◇ おわりに

以上、失敗 2 例と成功 1 例を紹介した。これらの事例から、Wiki を利用してプロジェクト研究を効率よく進めるためには、おおむね以下のような条件が必要であると思われる。

- Wiki の使い方を熟知した人間が、少なくとも一人はいること。初めの数ページで、他のメンバーに Wiki の使い方を納得させられなければ、ほとんど使われないまま放置される。
- 目的と手段が合っていること。主として Word の文書を、議論しながら叩くのであ

れば、Wikiの出る幕はない。

- 規模が適当であること。WikiPedia[9]のように数千人、数万人が参加するWikiのプロジェクトもあり得るが、不特定多数の交流場所にするのでなければ、顔の見える範囲、せいぜい5、6人から多くても10数人までだろう。

Wikiは何を書き込んでもいいが、少なくとも、サイトの方向性を決める最初の数ページだけは、設置者(プロジェクトのリーダー等)が中心になって、十分検討しながら作っておく、どうやらこの辺りがWiki使いこなしのコツということになりそうだ。

注

[1] 本来WikiWikiWebという名前はWard Cunninghamが開発したオリジナルにつけられた名前であり、これをもとに設計された派生物はWikiエンジンまたはWikiクローンと呼んで区別している。本稿では両者をあわせてWikiと呼ぶことにする。

[2] 雪印広報担当者の言。ただし、雪印に対するプレゼンを実施した直後に、二つめの事件が発覚したため、本当に実効性があったか、確認できないが。

[3] このため、Wikiを個人やグループによる情報提供サイトのコンテンツ管理に用いることも多い。そのような用途に関しては本誌の別の記事を参照されたい。

[4] WikiNameのように、先頭を大文字にした単語二つを並べたもの。通常WikiNameと呼ばれる。日本語対応のWikiでは、[[ゐき]]のようにブラケット二つで挟んでも、ページタイトルとして認識するものが多い。

[5] 結城浩『結城浩のWiki入門』PP.3　インプレス　2004.3

[6] Yahoo! Groups（http://groups.yahoo.co.jp/）など。

[7] 最近のWikiは、共同作業だけではなく、コンテンツ提供にも使われるので、メンバー管理機能を備えているものも増えている。

[8] それならボランティアにやらすなよ、とも言える。

[9] Wikiの仕組みで百科事典を作ってしまおう、というプロジェクト。
http://ja.wikipedia.org/wiki/E3%83%A1%E3%82%A4%E3%83%B3%E3%83%9A%E3%83%BC%E3%82%B8

学術情報発信ツールとしてのWiki

千田　大介（ちだ　だいすけ）

◆ はじめに

Wikiや Blogの流行については、昨今、さまざまな雑誌などで取り上げられており、容易に他のサイトと連携できるオープン性と批評メディアとしての特質、サイト構築の簡便性などさまざまな理由が指摘されているが、WWWの利用が定着し情報の発信者・受容者ともに成熟した結果として、デザインに凝ったサイトよりも実質的な情報を発信するサイトへの評価が高まっていることも一つの要因なのではあるまいか。

そもそも学術サイトとは、大学の学科組織や研究会サイトであれ個人サイトであれ、第一に研究・教育に有用な情報を継続的に発信し続けることが求められる。この点において、オープン性と簡便性とを備えた情報発信ツールであるWikiは、まさしく学術サイトの構築に適しているといえよう。

かかる観点から、筆者は、各種Wikiクローンの多言語化を進めるとともに、個人サイト「電脳瓦崗寨」[1]およびWebマスターをつとめる中国都市芸能研究会サイトをWikiに移行させ、さまざまな学術情報の発信に活用している。小文はその報告である。

◆ Wikiの多言語・多漢字化

◉ YukiWikiとUniWiki

日本で作成・配布されているWikiクローンの大半は、日本語（Shift JISやEUC-JP）にしか対応していない。このため、Wikiを中国学用途に用いるにあたっては、多言語・多漢字処理というくつもの問題をクリアしなくてはならない。

日本語版のWikiクローンは複数存在するが、Perlによって記述されたものとしては、YukiWiki[2]が最も広く利用されている。しかし、YukiWikiはjcode.plを利用する日本語CGIであり、多言語・多漢字ドキュメントを扱うのは難しい。そのため、筆者がYukiWiki 2.05をUTF-8化し多言語処理に対応させたのがUniWikiである。最新版はβ 1.56で、筆者のサイトで配布している。

実のところ、日本語のPerl CGIをUTF-8による多言語CGIに改造するのは比較的簡単である。文字コードを変換し、HTMLヘッダ生成部分とjcode.pl関連部分を調整するだけでよい。スクリプト中に正規表現による置換部分がある場合は、左辺にASCIIコードの文字しか使えないという制限はあるものの、最新のPerl5.8のみならず、Perl 5.xで問題なく利用できる。

UniWikiでは、多言語混在ページに対応するため、文字列の言語を

```
<span lang="xxx">
```

タグで指定できるよう整形ルールを追加している。最新のブラウザでは、これによって表示フォントが当該言語のものに自動で切り替わるので、Unicodeで統合されている「骨」と「骨」、「差」

と「差」などの日中の字形差も表現することができる[3]。

●縦ゐき

本誌創刊号でもレビューされているように、Internet Explorer 5.5 以降は縦書き HTML の表示に対応している[4]。しかし、それを応用した縦書き Web ページは非常に少ない。かなり以前から、テーブルを利用した縦書き Web サイトや、横書きの文字を並べ替えて縦書き表示する CGI は存在しているので[5]、需要はそれなりにあるものと思われる。それにも関わらず縦書き HTML が普及しない原因は、ブラウザの互換性問題もさることながら、縦書き HTML に存在するさまざまな制限や癖を解説したマニュアルが乏しく縦書きが可能なことすら余り知られていない[6]、縦書きに対応した HTML エディタがない、といった点にあると思われる。

かかる縦書き HTML につきまとう困難の解決を目指して筆者が開発したのが「縦ゐき」である。縦ゐきは、YukiWiki・UniWiki をベースに、HTML 生成部分とスタイルシートとを調整して、縦書き Web ページの自然な表現を実現したものである。執筆時点での最新版は β 1.12、筆者のサイトで配布している。動作環境等は UniWiki と同じである。

開発にあたり、縦書き Web ページを必要とする具体的対象として、漢文の訓点文を想定した。漢文独特の返り点の表現や、再読文字の左右に付されるふりがな・送りがな等は、縦書きでなくては表現しえない。HTML でそれらを表現するのは非常に困難であり、Web での表現は画像・PDF などに頼っているのが現状である。このことが漢文訓読文献のデジタル化を進める上での一つの障害となっていた。

縦ゐきは、縦書き HTML とテーブルの組み合わせによって、整形ルールに従って記述されたテキストを訓点文として表示する機能を備えている。テーブルを用いているため、HTML 表示からワープロソフト等へのコピーはできない。しかし Wiki はデータをマークアップテキストとして保存するで、そのテキストデータを介してデータの共有・交換が容易に実現できるという利点がある[7]。

将来的に XHTML 1.1 に見える ruby 機能を実装するブラウザが出現し、縦書き＋左右のルビが実現された場合でも、縦ゐきは訓点文生成部分を修正するだけで対応できるので、もとのデータが無駄になることはない。

編集(管理者用) | 編集 | 差分 | 新規作成 | 一覧 | RSS | FrontPage | 検索 | 更新履歴

多言語混在例

- 1.中国語教材
  - 言語指定しない場合
  - 言語指定すると
- 2.漢字の地域差比較
- 3.アラビア語の混在処理
- 4.ピンイン指定

1.中国語教材

言語指定しない場合

今天到什么地方去?	Jīntiān dào shénme dìfang qù?
今日はどこに行きますか？	
看着电视吃饭。	Kànzhe diànshì chīfàn。
テレビを見ながら食事する。	

簡体字「么」だけが中国語フォントで表示されますが、その他の文字は日本語フォントで表示されるので、見かけが不揃いになります。下の例では、それがより一層顕著に表れているとともに、「看着」が日本字形になっており、縦棒が連続する中国語字形になりません。

言語指定すると

今天到什么地方去?	Jīntiān dào shénme dìfang qù?
今日はどこに行きますか？	Jīntiān dào shénme dìfang qù?
看着电视吃饭。	Kànzhe diànshì chīfàn。
テレビを見ながら食事する。	Kànzhe diànshì chīfàn。

図 1：UniWiki での多言語混在表示例

⦿ PukiWiki の UTF-8 化

PukiWiki は、PHP で記述された国産 Wiki クローンである。メニューバーを備えた洗練されたインターフェス、プラグインによる豊富な機能と優れた拡張性を備えており、筆者もサイトの運営に利用している。執筆時点での最新版は 1.4.3[8]。

PukiWiki には、UTF-8 化による台湾版が存在しているので[9]、それをベースに、HTML ヘッダ等の言語設定部分を書き換え、日本語スキンファイル等を UTF-8 にコンバートすれば、比較的容易に多言語化が実現できる。また、default.

ini.php に整形ルールの拡張設定が用意されているので、span タグによる言語指定なども簡単に追加できる [10]。

ただし、PukiWiki は PHP 4.1.0 以上でないと動作しないため、設置できないサーバも多い。その場合は、Perl ベースの CGI を利用するしかなかろう。

◇ Wiki による研究会サイトの構築

⦿ 中国都市芸能研究会 PukiWiki

中国都市芸能研究会は、近現代中国の都市における芸能とそれをとりまく事象について研究する、分野横断的な研究会であり、会誌『中国都市芸能研究』を毎年発行している [11]。

中国都市芸能研究会サイトは、執筆時点では、三つの PukiWiki から成り立っている。第一は一般向けの情報公開用 Wiki[12]。第二は焦循『劇説』Wiki[13] で、会読成果を公開するとともに、後述するように古典戯曲総合データベースとして発展させていくことを計画している。第三は対外非公開の会員専用 Wiki で、整理中の各種資料や、名簿などの非公開情報・個人情報などを交換する必要から設置している。

⦿ 研究会サイト Wiki 化のメリット

研究会サイトを Wiki で運営することの最大のメリットは、いうまでもなくサイト更新の負担を分散できることにある。

FTP によるサイト更新では、パスワード管理が不可欠である上に、HTML・FTP 等の知識が欠かせないので、勢い数少ない電脳派の人が Web マスターに指名され、更新負担が集中してしまうことになる。一般的に、人文学・中国学の領域では電脳人材が豊富とはいえず、事務作業の電脳化が進んだ昨今は、限られた電脳派の人に仕事が集中する傾向に拍車がかかっているようである。このため、忙しさからサイトの更新頻度が低下する、更には Web マスターの異動によってサイトが死んでしまう、などという事態がしばしば発生する。

現実的には、研究会・学科等を問わず、Web を通じた情報の公開に理解のある人はごく少数であり、いかに Wiki を利用したところで、だれもが積極的にサイトの更新や情報公開を始めるとは考えられない。しかし、Wiki は HTML や FTP の知識を必要としないので、インターネットに接続し BBS を利用できる程度の PC スキルさえあれば、誰でも簡単にページの更新や追加ができる。それゆえに、Web マスターは具体的なページ作成・更新作業を比較的気楽に会

図2：中国都市芸能研究会　PukiWiki

員に割りふることができ、サイトメンテナンスの負担を分散することができるのである。ひいては組織全体の PC スキルや情報公開意識を向上させることにも繋がろう。

また、PukiWiki ではファイルをページに添付してアップロードすることができる。つまり、Web サイトを手軽にサーバの共有スペースとして利用することができるわけである。この機能を利用することで、写真や動画、Word・Excel のファイル等を共有することができる。

中国都市芸能研究会サイトでは、主に会員専用 Wiki でこの機能を活用している。全国に会員が散らばる研究会では、研究例会等の欠席者との情報共有化が一つの問題となる。欠席者にレジュメを一々郵送するのは非常な手間で、実現が困難であったが、PukiWiki の添付ファイル機能を利用することで、基本的にこの問題は解決された [14]。このほか、現地調査の記録写真、校正中のテキストデータ、さらには各種申請書の共有化・交換等にも利用している。

以上は、Wiki のコラボレーションツールとしての特性を生かした利用方法といえよう。

⊙ 研究論文の公開

● PukiWiki による人文学論文の処理

本誌第三号の記事中で、『中国都市芸能研究』PDF 版のデータ容量が大きいという問題を指摘したが、論文データの公開に PukiWiki を用いることで、その問題の解決を試みた。

前述のように、PukiWiki は非常に機能が豊富であるため、人文学の学術論文に用いられる書式であれば、ほぼ問題なく表現できる [15]。例えば、ページに画像を添付してアップロードするとともに、任意の位置に読み込むことができる。表組みでは上下・左右のセルを結合させることもできる。また、丸カッコ二つで文字列を囲うことで、脚注を追加することができる。注釈番号がアラビア数字に限られるなどの制限はあるものの、実用上問題ないといってよかろう。

ただ問題となるのが、見出しのレベルである。一般に国産 Wiki クローンでは、ページ中の子見出しのレベルは三段階まで（h2 ～ h4）の対応である。しかし、多少長めの人文学論文であれば、四段階、あるいは五段階を越える見出しレベルが使われることも珍しくはない。

この問題を解決するためには、PukiWiki そのものを改修する必要がある。具体的には、アスタリスクから始まる行を見出しタグに置換する箇所、html.php の 316・323 行、および convert_html.php の 168 行に見える「3」を任意の数字に書き換えればよい。例えば「5」に書き換えれば、見出しを五段階（h2 ～ h6）まで利用できるようになる。

中国都市芸能研究会の刊行物のうち、科研費報告書は Word で、『中国都市芸能研究』は Adobe InDesign で組版している [16]。Word ファイルは、独自のテンプレートを作成し、スタイル機能を用いて作成したものである。従って、Word 上で書式を検索して見出しや引用箇所の行頭に記号を挿入することで、Wiki の整形ルールに従ったテキストに変換することができる。InDesign の場合は、データを独自タグ付き UTF-8 テキストで書き出すことができるので、それらのタグを置換すればよい。

脚注はもとの InDesign データでは本文とは独立したストーリーに配置されていたが、さしたる量でもなかったため、二重括弧を付けて本文に埋め込む作業は手で行った。

Wiki は、ディレクトリや階層構造を持たないために、作成したページの整理やアクセスにとかく困難がつきまとうのだが、PukiWiki ではその対策としてページ名の階層処理と、サブメニューという機能が用意されている。ページ名の階層処理は、「分類名 / ページ名」と命名することで、ページをカテゴリ分類する機能である。階層は何段階にも分類することができる。サブメニューは、「分類名 /SubMenu」というページを作成し、「MenuBar」に「#includesubmenu」と記述することで、分類名が含まれるページを開いたときに、メニューバーにサブメニューが表示される機能である [17]。

論文データの公開に際しては、この機能を活用し、「『都市芸研』第一輯 / 論文タイトル」というページ名でページを作成し、サブメニューとの連繋によってアクセシビリティーの向上をはかった。

●利点と問題点

人文学の学術論文の公開では、PDF、あるいはWord などのワープロソフトからエクスポートした HTML が使われることが多い。しかし、前者にはファイルサイズが大きくなるという問題が、後者には独自のタグが大量に付されてしまいアクセシビリティーの高いシンプルな HTML の生成が困難であるという問題がそれぞれ存在する。

Wiki では、Word や DTP ソフトのスタイルを検索し置換・整形することで、比較的簡単に Wiki データを作成することができる。しかも、Wiki に登録してしまえば常にシンプルかつ正確な HTML が生成されるので、HTML に比べて作業の手間をかなり低減できる。全文検索機能が利用できるのも嬉しい。ただし、Wiki ページが増えるに従い、検索速度が遅くなるという問題はある。

学術研究では常に研究が深化し新たな資料が発見されつづける以上、研究論文にも本当の意味での完成はありえないし、いかに万全を期しても、それでも誤りや校正漏れが発生することは避けられない。この点、Wiki による公開には、論文を容易に修正できるというメリットがある。

ただし、学術論文のデータは、それぞれの著者の研究業績であり、改訂・書き換えは著者の責任のもとに行われなくてはならないので、不特定多数による書き換えが可能な状態で公開するのには、些か危険がともなう。このため論攷に関しては、部外者が更新できないよう、パスワードによるロックをかけている。

ページのロックを解除し、ページに閲覧者が掲示版の感覚でコメントを書き込めるように一行コメントを付け加えるなどすれば、閲覧者とのオープンな対話によって論文を更新していく、つまり知的生産の場として Wiki を活用することも可能になろう。もっとも、Wiki には悪意ある書き込みや改竄・ページ消去などのリスクがどうしてもつきまとうので、運用方法には工夫が必要となる。また、論文は引用されるものであることを考えれば、頻繁な書き換えに適さない面もあるし、書き換えが数次にわたる場合のバージョン管理問題もある。この点は、今後の課題である。

⦿ 電子テキストと会読成果の公開

図 3：『劇説』Wiki

●『劇説』Wiki

『劇説』は、清代中期の人、焦循の手になる全六巻の筆記小説である。歴代の戯曲・演劇の作品・作家と上演に関する記事を編纂したもので、中国古典戯曲を研究する上で最も重要かつ根本的な資料であり、中国古典戯曲史の大半は同書をもとに記述されているといっても過言でないほどである。

『劇説』の排印本は数種類刊行されているが、しかしながら、引用文献の原典を一々確認して校訂を加えたものは存在しない。そこで、中国都市芸能研究会では、同書を会読して校注を施すとともに邦訳をすすめており、その成果を逐次、「『劇説』

校注　附：訳注稿」として『中国都市芸能研究』誌上に発表している。焦循『劇説』Wiki は、それを PukiWiki 上に公開したものである。

● Wiki データファイルへの加工

『劇説』Wiki は InDesign で DTP したデータを元にしたものであり、作成しなくてはならない Wiki ページは、本文・訳文・注釈あわせて数百ページにも及ぶ。このため、データを手作業で一々 Wiki に登録するのは現実的ではない。

多くの国産 Wiki クローンでは、各ページのデータはテキスト形式で保存され、ファイル名はタイトルの文字列を文字コード番号に変換したものになる。例えば、UTF-8 を使用している場合、「元曲」というタイトルのファイル名は「E58583E69BB2.txt」になる。このファイル名は、Perl スクリプト等を用いて容易に生成することができる。

さて、紙版の『劇説』校注では、校訂済み原文に中国語の注釈を、校注・戯曲用語注・作品名注・人名注などに分類して付し、その後に日本語訳と訳注とを掲載する形になっている。注釈を内容ごとに分類したのは、歴代の戯曲資料をひろく収集した『劇説』の全編に施した注釈を抽出・加工することで、簡明な古典戯曲辞典や用語集を作成しうるものとの見通しを持っており、将来的な再編集の便を考慮したものである。

作業は、InDesign からエクスポートした InDesign タグ付きテキストデータをもとに、まず EmEditor[18] を用いて不要な InDesign タグの除去などの加工を施した。その上で本文・注釈・訳文などを、それぞれ別ファイルに抽出した。

注釈のうち校注と訳注とは、それぞれ原文・日本語通釈に脚注として埋め込んだ。その他の注釈は、リンクの便を考え、それぞれ独立したページに加工した。まず見出し語が重複するものの内容を調整し、他の脚注を参照するだけのものを削除した上で、Perl を用いて見出し語ごと、つまりは一行ごとに別ファイルに切り出し、Wiki データを生成した。注釈は、本文中から容易にリンクできるよう、それぞれの見出し語をファイル名とした。

本文と訳文は、校注・訳注を脚注として手作業で埋め込み、やはり Perl を用いて記事ごとに別ファイルに切り出した。ファイル名は、本文については記事の通し番号、訳文については「通し番号 / 通釈」とし、対応する本文と訳文との間を容易に行き来できるよう、それぞれのページへのリンクを付け加えた。日本語 Wiki クローンでは、一般に二重のブラケット（[[]]）で囲んだ字句には、その字句をタイトルとするページへのリンクが設定されるので、リンクの設定は容易である。それを利用し、本文・訳文に現れる注釈の見出し語についても、各注釈にリンクするよう一括置換した。

こうして作成されたデータを PukiWiki のデータディレクトリにアップロードし、パーミッションを変更してユーザーにも書き込み権限を与え、『劇説』Wiki の公開作業は完了した。

●ハイパーテキスト化の方法としての Wiki

『劇説』Wiki では、メニューバーの目次に従って各記事を閲覧する。表示した記事本文中のリンクをクリックすれば、注釈にジャンプできる。また、多くの Wiki クローンはページタイトルをクリックすることで、そのタイトルをキーワードとして Wiki ページを全文検索する機能があるので、注釈の立てられた語彙の出現箇所を一覧しジャンプすることもできる。

このように『劇説』Wiki は、ハイパーリンク等の機能によってテキストのリニアな構造を解体した典型的なハイパーテキストとして成立している。ハイパーテキストは、WWW の登場とともにその革新的意義が盛んに喧伝されてきたが、中国学の分野では、コーパスこそ構築が進んでいるものの、学術的に有用なハイパーテキストはほとんど存在しないのが実情である。その理由は、HTML ファイルの本文に一々ハイパーリンクを張り込むのは手間がかかるし、また漢字のファイル名が利用できないのでファイル名を一々考えるかあるいはスクリプトで一括処理するしかなく、ある程度のレベルの電脳スキルが必要になってしまう、という点に求められよう。

『劇説』Wikiでは、既存のテキストに基づいて作成したため、Perlスクリプトによる一括処理でWikiデータファイルを生成した。しかし、会読の成果を整理・公開する場としてWikiを活用し、徐々にコンテンツを構築していくのであれば、二重ブラケットで本文の語句を囲い、注釈の見出し語をファイル名にしてページを作成すればいいので、さほどスキルの高くないユーザーであっても十分に対応できる。ゆえに、学術・教育に役立つハイパーテキストサイトを容易に構築しうるツールとして、Wikiは非常に有用であるといえよう。

●今後の拡張計画と問題点

前述のように、『劇説』校注は、将来的に中国古典戯曲・演劇の小辞典として加工することを目指している。これは、Wikiに注釈語彙の索引を設置することで、ほぼ実現でききのではないかとの見通しを持っている。

現在、『劇説』Wikiには、中国古典小説研究会が書同文公司に委託入力した中国古典戯曲作品の解題目録、『曲海総目提要』を仮に登録してある。『曲海総目提要』は、校正や異体字の統一作業が完了していないし、リンクの埋め込みも完全ではない。しかし、まがりなりにも『劇説』と『曲海総目提要』という古典戯曲研究における重要資料がWiki上で有機的に結合されているので、例えば『劇説』の本文上の「王實甫」リンクをクリックして注釈に移動し、注釈本文からさらに「西廂記」注釈に跳び、さらに「西廂記」の含まれる全てのページを検索する、というような使い方が可能なのであるから、現状でも十分、研究・教育に利用しうるものと思われる。

今後、『劇説』校注の作成を進めるとともに、さまざまな古典戯曲資料や各種目録などのデータをWikiに登録し、相互に連携させることで、総合的な古典戯曲研究データベースとして発展させていく計画である。

ただ、前にもふれたように、PukiWikiやYukiWikiでは、ページ数が増えるにつれてページの表示や検索の速度が極端に低下するという問題がある。今後、WikiからHTMLファイルを生成させる、あるいはSQLを用いたサイト管理システムを導入するなどの対策が必要になるかもしれない。

また、『劇説』では本文は繁体字、訳文は常用漢字を採用するため、本文と訳文との連携が必ずしも十分ではない。異体字や類義語の一括検索の実現も、今後の課題である。

◆ 終わりに

以上のように、学術情報発信にWikiを活用する試みは、現時点では一定の成果を上げている。従来から指摘されているコラボレーションツールとしての適性もさることながら、ハイパーテキスト構築ツールとしても有用であることは、その人文学分野における応用の可能性を示している。

本誌レビューでもしばしば指摘されているように、国内研究・教育組織あるいは個人のWWWを通じた情報発信は全体としてはまだまだ立ち後れている。そのような状況を緩和する手段として、Wikiは一つの有力な選択肢となろう。

注

[1] http://wagang.econ.hc.keio.ac.jp/

[2] 結城浩氏作。http://www.hyuki.com/。執筆時点での最新バージョンは2.1.2。

[3] 本誌第3号「WWWブラウザ」レビュー（師茂樹）参照。

[4] 「ブラウザ」レビュー（師茂樹）。

[5] それらは、テキストそのものの排列を入れ替えて縦書きを実現するため、テキストのコピー・検索ができないという問題もある。

[6] たとえば、ただ記述方向を縦書きに指定しただけでは、一行十文字程度の横長のページになってしまう、といった問題がある。縦書きHTMLについては、縦ルビ（http://www012.upp.so-net.ne.jp/tateruby/）の解説が参考になる。

[7] 縦ゐきの訓点置換ルールは、筆者が配布しているWord用訓点マクロのそれと互換性が無い。訓点マクロを改修し、縦ゐきとワープロソフトとのスムースな

データ交換を実現することは、今後の課題である。

[8] http://pukiwiki.org/

[9] http://pukiwiki.kandix.idv.tw/

[10] 改造の具体的手順については、筆者のサイト（注２参照）の「PukiWiki 改造覚書」を参照して頂きたい。

[11] 本誌第３号拙文「電子版学術雑誌をめぐる諸問題」参照。

[12] http://wagang.econ.hc.keio.ac.jp/~chengyan/

[13] http://wagang.econ.hc.keio.ac.jp/~chengyan/jushuo/

[14] ただし、文献コピー資料をスキャンしてアップする必要がある、という問題はのこる。

[15] PukiWiki の人文学論文記述への適正については、PukiWiki.org「Use PukiWiki/ 論文だって書けちゃう」ページ参照。

[16] InDesign の多言語・多漢字組版機能については、本号およびバックナンバーの InDesign レビュー参照。

[17] 以上の機能については、PukiWiki.org のプラグイン解説参照。

[18] EmEditor については、本誌第４号「テキストエディタ」レビュー（山田崇仁）参照。なお、バージョン４で UTF-8 の正規表現検索・置換・Grep 等に対応した。入手先は http://www.emeditor.com/jp/

授業に生かす Weblog と UniWiki

その特性と活用

小川　利康（おがわ　としやす）

◇ はじめに

筆者は2003年秋学期からWeblog（Movable Type Ver.2.6x）[1]を利用した授業用ウェブサイトを公開している[2]。これは03年春学期に田邊鉄氏の主宰するウェブサイトStudy Chinese[3]に参加、PukiWikiを授業で利用した体験が契機となって始めたものである。

PukiWiki・Weblogに続き、04年春学期からはUniWiki（β1.56千田大介氏作）も併用、試行錯誤のなかでそれぞれの特性も実感的に理解することが出来た。これまで一年半の利用経験をもとに、中国語教育への活用方法を書いてみたい。

⦿「ブログ」「はてな」人気の理由

WeblogとWikiの歴史をここで紹介するゆとりはないので、別途書籍やネット上での議論をご参照いただくとして[4]、日本におけるWiki, Weblog流行の主たる理由がウェブ管理の手間を大幅に軽減した点にあることは間違いない。従来は、全体のページ構成・デザイン、HTMLコーディングに至るまでの編集作業から、ファイルftp転送に至るまで、幾つもの工程があり、ある程度自動化できるにせよ、やはり煩瑣だった。ところがWikiやWeblogを利用すれば、その手間は相当軽減されるうえ、ネット環境とブラウザさえあれば、いつでもどこでもオンラインで更新できる。この手軽さゆえに、大手プロバイダが競ってWeblogの無料提供を宣伝している[5]。HTMLファイルで一つ一つページを作るのは決して容易ではないが、Weblogなら器は全て用意されているので、必要なのは盛り込む文章だけである。「はてなダイアリー」[6]（ソフトウェアとしてはWikiに近い）にしても同様の理由から人気を博しているものと思われる。

この管理の省力化は当然のことながら授業用ウェブサイトを運営するうえでも大きなメリットである。筆者がWeblog、Wikiを導入した最大の理由もこの点にあった。だがCMS（Contents Management System）導入の意味は省力化に止まるものではなかったのである。

◇ 教育向けCMSの登場

HTMLを手作業で作成する方法に対し、CGIなどを経由して動的にHTMLファイルを生成するシステムを利用するメリットは省力化だけではなく、例えば、入力データを自動的にRSS（RDF Site Summary）で提供したり、時系列・カテゴリ別HTMLファイルを生成したり、ハイパーリンクを自動的に施すなどの機能を持つ。従来ならば手作業に頼っていたウェブページ間の整合性・連携の維持が自動化できるのも大きなメリットで

ある。

このような仕組みを Web コンテンツ管理システム、あるいは CMS と総称している。例えば、大規模なコミュニティ運営に適した XOOPS[7] や教育用に特化した Moodle[8] などがその代表であり、いずれも GPL に基づき無料で提供されている。

いずれも無料とはいえ、極めて多くの機能を備え、ユーザの必要に応じて新たなアドインソフトを開発することも可能で、拡張性にも優れている。例えば、教育向けに特化された Moodle ならば、標準インストールの状態のままでも、

1) 科目、コース名設置
2) 通知掲示板設置
3) ユーザごとに利用権限を設定
4) フォーラム、チャット開設
5) 小テスト作成、回収
6) レポート回収

といった多彩な機能が直ちに利用できる。

本誌前号でレビューのあった商用版の Blackboard Learning System（以下 BLS）を彷彿とさせるシステムだが、PHP4.1.0 以降インストール済みのサーバで、なおかつ MySQL が利用可能なら、一時間足らずで簡単にインストールできる。現在、英語版のほか、日本語（EUC-JP）版、簡体字版、繁体字版、ハングルなど二十数カ国語のバージョンが提供されているが、UTF 化されていないため、多言語混在には若干問題がある [9]。

もしも完全な e-learning（正確には WBT=Web Based Training と呼ぶべきかも知れないが）として授業ウェブを開設するのであれば、Moodle などの大規模システムは恐らく必須とも言えるだろう。実際これだけ多機能なシステムが無償で提供されているのは驚くべきことである。

確かに CMS としての Moodle の秀逸さは疑うべくもない。UTF への対応が行われてないこと、個別のフォーラムやチャットの機能がやや貧弱であるなど、欠点はあるものの、これだけの多機能をシームレスに管理できるのは魅力である。受講生全員にユニークな ID を付与し、それぞれ資格に応じたアクセスコントロールも行えるので、オンラインだけで採点し、成績判定を行うことも十分可能だ。

⦿ 授業を補完するシステムを

だが、現実の学校教育でこのような CMS がどこまで必要とされるのかは未知数である。今のところは教室での対面授業が主流であり、ここまで授業内容をオンライン化するメリットはない。例えば、オンラインで小テストや課題を学生に与えるよりも、授業中に学生に手書きで答案を書かせ、自己採点をした方が手間も省けるうえ、学習効果も高く、学生達自身も手書きによるテストの方を圧倒的に支持している。

従って、現時点で Moodle の豊富な機能を必要とするのは職業教育や資格教育など、対費用効果がシビアに測られる教育分野に限られるだろうし、そのシステムに対応した人員配置が必須とな

図 1：Moodle のトップページ。標準では三カラム構造で中央に会議室の内容が掲示される

る。WBT自体そもそも時間・経費節約のために生まれたシステムである以上、当然のことかもしれないが、現在大学で学ぶ学生はもっと基礎的なトレーニングを必要としており、キーボードとモニタだけで学習活動を完結させることは難しい。

では、現在の教育体制のなかでCMSは全く不要かというと、そうともいえない。

教室での対面授業を中心とした学習活動はオンライン教育より有効だが、教室における授業形態も万能ではなく、多くの制約を受けている。ここでCMSの役割を考えるにあたっては、オンラインで何が出来るかを問う前に、対面授業で何が足りないかを問わねばならない。

そもそも授業というのは時間と空間を制約するところで成立している。決められた時間に同じ場所で受けるからこそ、授業が成立する。そのデメリットも否定できないが、何よりも強い拘束こそが学習活動を支える大きな仕組みとなっていることに注意しなければならない。オンラインで公開されている無数の学習コンテンツや図書館に所蔵される万巻の書があれば、授業はそもそも必要ないにもかかわらず学校教育が必要とされる理由の一つは授業という「場」の持つ「磁力」に他ならない。

「場の磁力」は単なる義務感でなく、学習意欲、人間関係（対友人・対教員）など幾つもの要素から醸成され、教材以外の要素からも強く影響されながら学んでいる。これがオンライン教育によって最も代替しにくい部分である。

むろん負の側面も免れない。限られた時空のなかで展開されるために、授業で伝達される情報はしばしば教員側からの一方通行になりがちで、情報のほとんどは音声に依存するため、終了後はほとんど残らない。そのためにこそ教科書などの紙媒体があるわけだが、毎回の授業内容を忠実にトレースしたものではない。

現状の授業に最も必要なのは恐らく、毎回の授業内容を分かりやすく整理し、跡づけることができて、学生からのフィードバックを受け付けるシステムではないかと思われる。

⊙ 中国語教材を扱うためのUnicode対応

中国語もしくは中国関連の授業に関するコンテンツであるから、中国語（簡体字、繁体字）がナチュラルに扱えるシステムでなければならないことは言うまでもない。以下では簡単にUnicode（UTF-8）への対応状況を見ておこう。

● Movable Type（MT）2.661

開発元Six Apartが無償提供する英語版をベースに、有志提供の日本語化パッチ[10]をあてることで、日本語EUC版もしくは日本語UTF-8版を生成できる。このバージョンについては新バージョンと異なって、基本的にどのような利用形態でも無料で使える。英語版のまま使うのであれば、そのままUTF-8として利用可能。あとはサーバ側がPerl5.0以降であれば問題ないようで、ほとんどのサーバが対応している。現在も使い続けるユーザが多い。

● Movable Type（MT）3.01D

開発元Six Apartが定めるライセンスに基づき有償で提供され[11]、日本語版も開発元から提供されるようになった。新たなバージョンアップにより追加された機能はコメントスパム対策としての認証機能だけで、期待されたMoblog機能[12]は含まれていない。商用版とするための仕切り直しバージョンといえる。

日本版はmt.cfgの設定によりUTF（標準）のほか、Shift-JIS, EUCから選択できる。だが、ここで日本語（UTF）版を選択しても、テキスト入力画面で中国語が正常に表示できず、文字化けしてしまう。実際に出力されるHTMLファイルでは簡体字も表示されるので、とりあえず別途エディタで編集のうえ、入力画面に貼り付けても良いのだが、これでは余りに不便である。

この現象を回避するには日本語版のスクリプトを自ら修正するか（今回より有償版となったため余り推奨できない）、英語版を選択しなければならないため、アップデートのメリットはない。旧バージョンの利用を強く推奨する。

このほかにも Weblog に分類されるソフトウェアについても幾つか試してみた限りでは類似した問題を回避することは難しいようで、必ずしも MT より Unicode 対応が進んでいるとはいえない。たとえば Nucleus[13] はテキスト入力の画面で文字化けして中国語が表示できないものの、実際に HTML に出力してみると、問題なく表示される。スクリプトをある程度修正すれば問題は解決可能かも知れないが、MT3.0（日本語版）と同様の問題を抱えている。文字化け回避のために英語版を用いても、言語指定が Latin と固定されているため、中国語ばかりか、日本語も正常に表示できなくなる。p_blog[14] は出力される HTML こそ UTF だが、内部のスクリプトは EUC で動作しているため、簡体字は全く受け付けない。なお、現在 @nifty などが提供する Weblog サービスも基本的に MT3.0 に近い仕様なので、入力画面で簡体字は文字化けする（出力は正常に行われる）。

従って、現状では MT の旧版が最も使いやすく、旧版が手に入らない場合でも MT3.0（英語版）あるいは Nucleus が推奨される選択肢となる。Nucleus は総じて動作が軽快であり、今後の進化が期待される。

● PukiWiki

多数の Wiki Engines のなかでも、PukiWiki は現在最も人気のあるクローンと思われるが、標準のエンコードが EUC で、多言語処理に未対応であるため、PHP スクリプトを書き換えなければならず、いささかスキルを要する[15]。

この Wiki の UTF 化を行い、はじめて中国語教育に応用したのが田邊鉄氏であり、「ゐきゐきコラボレーション授業」（中国語 CAI 研究会総会発表 2002 年 12 月 8 日）を皮切りに、中国語教育だけでなく共同研究への活用まで幅広く取り組み、現在も PukiWiki 1.3（改造版）を利用した中国語学習ポータルサイト「ちゃいなび」[16] を展開している。

● UniWiki

じつは中国語教育に Wiki・Weblog を利用するには単に UTF に対応させただけでは、日中韓の異体字区別まで正確に行えない。このため、正確な字体を表示できる工夫が必要だった。

図 2：MT3（英語版）の管理画面

その解決策として開発されたのが千田大介氏による UniWiki である。YukiWiki をベースに開発され、簡体字・繁体字・日本語・ハングルの字体を区別するための整形ルールを追加し、機能面では PukiWiki ほど豊富ではないものの、必要充分な機能を備えている。千田氏はこのほか漢文の訓点付き縦書き表示を簡便に行うための「縦ゐき」開発も進め、研究面への活用を進めておられる。こちらも詳しくは本誌掲載の千田氏自身の論考を参照して欲しい。

以上のように、Perl などのスクリプト側でも Unicode 対応が進み、先行する諸氏の努力の恩恵により、MT・Wiki どちらを使っても、おおむね問題なく中国語が扱えるようになっている。

◇ Movable Type：運用の実際

◉ Movable Type と Wiki の特性の違い

さて具体的な活用の実際を見てゆこう。

Weblog とは平たく言えば Web 日記のことである。従って、データは時系列で配置される。時間は設定により、日ごと、週ごと、月ごとにアー

カイブを自動作成してくれる。この時間を軸としたデータ構造がWeblogの基本である。授業もスケジュールに従って決まった期間展開されるものなので、Weblogとの相性は非常によい。閲覧する学生側からしても必要とするデータを時間軸に沿って遡及しながら探してゆく方が見落としもなく合理的である。

とはいえ、異なる科目の内容が混在していては見にくくなる。このためWeblogのCategoryで分類する。Individualという項目で投稿者別にアーカイブを自動作成することも可能で、複数の教員で運営する場合には有効だが、その際はWeblogそのものを複数立ち上げた方が合理的である。Wikiと異なり、Weblogは基本的に一人か少数のメンバーによって書き継がれてゆくのが前提のため、一つのシステムファイルのもとに、外見上は全く別々のWeblogを複数設置することが可能である。ユーザごとに異なるIDでログインすれば、複数の教員のWeblogを運営できる。

このように時間軸と科目別に自動作成されたアーカイブがWeblogの実体である。自動作成の際にはユーザが指定したテンプレート、スタイルシートに基づいてHTMLファイルが作成される。このためWeblog上でデータ入力をする際には、コーディングをほとんど意識しなくて良い。HTMLファイルとして出力されるまでの流れを見ると、

テキスト入力（テキストのみ）
テンプレート読み込み
　（ページレイアウト、文字スタイル指定）
HTML生成
　（カテゴリ別、月日別など複数生成）

という流れになる。レイアウトに必要なHTMLタグなどはテンプレートに入っていて、自動的に追加される。従って、設計思想からすると、テキスト入力の際にHTMLタグを入れるのは余り推奨されないのだが、さすがに太文字やハイパーリンクを入力するボタンは用意されている。

筆者の場合、中国語簡体字・繁体字を正確に表示するために、あえて入力ボタンを拡張したほどであるが[17]、これはWeblogの設計思想からすると邪道であることは間違いない。

これに対し、Wikiでは豊富な整形ルールが提供され、その大半が文章を階層化する（H1～H3など）整形ルールであり、オンライン上で利用できるアウトラインプロセッサともいえるほど多くの整形ルールが提供され、HTMLタグが本来目指していた文書の構造化を簡便に実現できる。

Weblogの場合、HTMLタグによる文書の構造化は予め用意されたテンプレートで実現するのが前提であり、Wikiほどの柔軟性はない。ここに両者の設計思想の際だった違いを見て取ることができるだろう。

端的に言えばWikiにおいてはデータ構造を決定するのは基本的にユーザの発想に依存している。整形ルールを上手に使えば、荷物の大きさや形に合わせて、入れ物も自由に作れるシステムである。一方のWeblogではデータ構造を決定するのはシステムであり、時系列やカテゴリが構造を決める。いわば用意さ

図３：拼・赤・簡・繁・PICのアイコンは、自作追加したもの

れたタンスにデータを放り込んでゆくというスタイルである。ある程度までは融通が利くけれど、時として、靴下一足のために大きなタンスを一つ用意してしまうようなこともあるだろう。両者の優劣を決めることはできないが、この特性の違いを意識して、活用方法を考えなければならない。

◇ 講義Weblogの構成

Weblogのトップページは左右2カラム構造とし、左列には常時最新記事が三本掲載されるよう設定している。これは標準的なWeblogのスタイルである。左列は記事が追加されるたびに古いものから消えてゆく。日常的にチェックしていればこのトップページだけで用が足りる。右側は最新記事以外のアーカイブにもアクセスしやすくするため、様々なショートカットを用意した（下図参照)。上から順に「カレンダ」からは月日別記事にジャンプでき、「科目別記事一覧」はカテゴリごとに分類された記事へのリンクを提供し、前学期までの記事は別途「(旧) 教材倉庫」からショートカットを提供する。このほかWeblog自身が提供する機能として、「サイト内検索」がある。

学期中は授業ごとにほぼ毎週更新することになるが、その内容は端的に言えば、これからやることの説明、終わったことについての評価にほかならない。これに授業開始に当たって配布したシラバスを加えれば、講義Weblogに含まれるべき要素のすべてになる。

- シラバス（授業計画進度予定）
- 事前：教材配布、授業及びテスト告知
- 事後：テスト解答、講評配布

具体的説明のために、今年度の「総合中国語」（三年次以上対象）を例に取ると、四月時点で配布するシラバスに春学期履修する内容（教材、日程、テスト日程）を掲げる。毎週更新する記事には

1) 小テスト用予習単語（簡体字とピンイン）
2) 授業内容の告知
3) 先週の小テストの解答とコメント

を掲載する。このほかにWeblogで取り上げたのは暗誦テスト課題の配布・説明、期末試験の出題方針を記事として掲載したものの、毎回の授業内容のまとめ等、授業そのものの記録を残すことまでは出来なかった。これは今後も物理的に時間が足りないので実現は難しい。

こうしたやり方は03年春学期当時に田邊鉄氏の主宰するStudy Chineseに参加して以来、大きく変わった点はない。授業の「補完」という位置づけに止まる限りは、これ以上できることは恐らくないのではないかと思われる。Weblogの機能として、記事に対してコメントをつけたり、ほかのWeblogからのトラックバック（参照リンク）を受け付ける機能があるものの、こうした双方向性コミュニケーションは対面授業での仕事である。Weblogでの交流がごく少数の利用に止まってい

図4：Weblogトップページ

るのはある意味では当然でもある。

⦿ UniWiki の構成

既に見て来たように Wiki の特性は文書データを柔軟に構造化できるシステムと参加者全員が自由平等に書き込みできる環境にある。「伽藍とバザール」[18] の喩えでいえば、間違いなくバザール方式を体現するシステムだ。その特性を生かした活用といえば、共同作業が中心となる授業が Wiki には相応しい。今年度から担当している中文電脳処理の授業は課題実習を通して PC リテラシーを学んでゆく授業なので、その授業向けの環境として、UniWiki を立ち上げることにした。実習のなかで作成したレポートはテキストベースなら、全て UniWiki に書き込んでもらい、お互いの作業結果が自由に閲覧できるようにした。

中国文献学 | 新規作成 | アップロード | 更新履歴 検索 | 一覧 | RSS |

中国文献学

欢迎! 歡迎! ようこそ！

このUniWikiは電脳瓦崗寨（http://wagang.econ.hc.keio.ac.jp/）の千田大介さん（慶應義塾大学）がYukiWiki?をベースに開発されたものです。中国文献学でHTMLの基礎を学ぶにあたり、ゼロからタグを勉強するのはあまりに煩瑣なので、構造を分かりやすく学ぶための手がかりとして、このUniWikiを利用します。

以下のUniWikiでも紹介したように、通常のホームページとは異なり、誰でもアクセスして、簡単に書き換えが出来る不思議なページです。Wiki入門をちょっと読めば、簡単にページが作れ、 Wikiの書き方でテキスト整形のコツがつかめれば、アウトラインプロセッサ風のアイディアノートとしても使えたり、グループで共有してアイディアを練ることも出来ます。また、千田さんの改造によって、UTF-8で動くWiki Engineなので、日中韓の文字を自由に入力でき、細かい字形も区別できます。凝ったデザインのページには向きませんが、シンプルなレイアウトは実質的で見やすいはずです。

受講生諸君は、まず以下の使い方を読んで、だいたいのやり方をつかみましょう。
６月よりUniwikiをアップデート！
こんな風に色や文字サイズも変更できるようになりました。！

UniWiki	UniWikiとはなにか？
Wiki入門	まずは読もう。簡単な使い方
Wikiの書き方	テキスト整形のルールを紹介します
練習用のページ	書き込み自由の練習ページ
中国文献学5	講義ノート：授業中に参照します（Unit3の続き）
講義Webloghttp://ogawat.daa.jp/cblog/	授業に関する重要な連絡

図 5：UniWiki の FrontPage

この授業で UniWiki が果たす役割は極めて大きい。コンピュータルームでの授業であるから、UniWiki は授業の事前事後の連絡だけではなく、授業内容そのものを担う存在でもある。

- シラバス（授業計画進度予定）
- 事前：授業に関する連絡、レポート告知
- 授業：講義ノート（教材提示）
- 事後：学生の課題提出及び掲示

UniWiki でも主役はやはり教員だが、Weblog と違って、学生からも自由にアクセスでき、指示に従って、課題報告を UniWiki に書き込んでゆくので、相対的には双方向性の交流に近くなる。今年度半年経過した段階での感触では、共同作業をする場所として、学生達は授業時間外にも UniWiki にアクセスする習慣は身につけてくれたと思う。

とはいえ、学生達の習熟度と授業課題が基本的には個人単位か、二名単位のグループで行われるため、UniWiki 上で学生間の双方向コミュニケーションが活発になったとは言えない状況である。春学期の反省点としては、やはり全員が共有する一つの目標なりゴールを設定する必要がありそうだ。どのような課題設定ならば、全員の積極的な参加が得られるか、まだ試行錯誤が必要だ。

◇ Weblog, Wiki の効用——まとめ

最後に Weblog・Wiki の果たした役割について総括しておきたい。

システム導入によって、従来よりも授業関連情報は質・量ともに向上したことは間違いない。これまでメーリングリストを利用して行ってきた情報伝達だけでは日本語に限られるために充分でなく、ウェブサイトでの授業情報提供もアドホックなもので、一定期間掲示した後は削除しており、更新も継続性に欠けていた。現在は簡単なメモを残す感覚で書いて、クリック一つでトップページに掲載できる。この省力化のメリットはやはり非常に大きい。第二に従来は更新頻度が低かったので、学生に何度も念を押さないと誰もわざわざウェブサイトを見てくれなかったが、事前に見ておかないと授業で困ることが分かれば、学生達

もまめにチェックするようになった。出来れば授業中に印刷して配って欲しいという要望もあるが、近年のブロードバンド普及の恩恵の賜物か、オンラインで見られる方が便利だという意見の方が多数を占めつつある。かつて思い描いていた状況に近づきつつあることは喜ばしい限りだ。第三に想定外のことながら、教材データが蓄積されるにつれて、教員自身が授業のあり方を振り返る機会が多くなった。正直なところ、オンラインで教材をほぼ全て公開するのは相当躊躇したが、実のところ一番熱心に読み返すのは教員自身である。過去のデータを参照するごとに読み返し、改善策を考えるというスタイルは授業改善の方法の一つと考えても良いと思う。

とはいえ、現状の Weblog にしても、Wiki にしてもオンラインでの双方向コミュニケーションはほとんど成立していない。「授業の補完」という位置づけであるからこそ、この結果に満足できたのだともいえる。

今後は対面授業の場でのコミュニケーションがオンラインへと広がってゆけば良いのだが、これはなかなか容易ではない。『伽藍とバザール』の喩えを今一度引くならば、

> バザールプロジェクトは、コーディネータやリーダの対人能力やコミュニケーション能力が優れていないとダメだ。(9 バザール方式の前提条件とは) [19]

「バザール」に集う学生達を増やすにはコーディネータとしての教師の力量が問われる。プロジェクト成功のカギは結局のところソフトではなく、人間なのだろう。当たり前だが改めて自戒としたい。

注

[1] Movable Type は Weblog（Web 日記）を作成するためのソフトウェアの一つでしかなく、注記３に記すとおり、Weblog のためのソフトウェアは他にも多数存在するので、決して Movable Type ＝ Weblog ではない。ただし、ここでは記述の便宜上、Movable Type での利用を前提とした Weblog について述べている。

[2] http://www.ogawat.net/cblog/
なお、04 年５月にリリースされた Movable Type3.0 は http://www.ogawat.net/blog/ で試験公開中。

[3] 現在は閉鎖。現在その理念は「ちゃいなび」に引き継がれている。

[4] Weblog については関連書籍としては、平田英治『Movable Type で今すぐできるウェブログ入門』（インプレス 2003 年８月刊）ホリコシ ヒデミほか『ウェブログ入門－ Blogger と Movable Type ではじめる』（翔泳社 2003 年７月刊)、松永英明『ウェブログ超入門！』（日本実業出版社 2004 年６月刊行）など。上記資料は Movable Type に限られるが、ほかにも古くは Blogger・blosxom から、最近は Nucleus・p_blog など多様なソフトウェアが存在する。従って、Movable Type の名前はあくまでもソフトウェアとしての名称であり、Weblog は注 [1] で述べたとおり、Web 日記という広範な概念である。
Wiki に関してはボウ ルーフ『Wiki Way―コラボレーションツール Wiki』（ソフトバンクパブリッシング 2002 年９月刊行）など。ネット上では、幾つかの議論や発言が重要と思われる。

Weblog の歴史

- 松永英明：日本のウェブログの歴史（詳細版）
  http://kotonoha.main.jp/weblog/000844_history.html
- 「ブログ草創期」論争
  http://bloggers.ja.bz/myojin/archives/000002.html

Wiki の歴史

- 九州産業大学電気工学科藤田研究室
  http://www.crow.dyndns.tv/pukiwiki/index.php

[5] 代表的なものとしては大手プロバイダとしてはじめて提供を始めたココログ（http://www.cocolog-nifty.com/）や excite ブログ（http://www.exblog.jp/）などが知られる。

[6] http://d.hatena.ne.jp/

[7] http://jp.xoops.org/ 参照。専用のインスーラーが附属し、最も扱いやすい CMS といえる。

[8] http://moodle.org/ から入手できる。日本では山口大学（共通教育英語部会）が比較的早く導入しており、日本語マニュアルも独自に提供している。http://ds21.cc.yamaguchi-u.ac.jp/~eigo/moodle/user/view.php?id=2&course=1 なお、筆者自身も Moodle の運用を試みたことがあるが、現在は既に撤去している。関連記事は http://www.ogawat.net/archives/000088.html

[9] Moodle では掲示板は全て RTF（Rich Text Format）での表示も選択できるので、ワープロ感覚での表現が可能である。Unicode 未対応ながら、中国語も RTF を利用すれば表示が可能になるので、回避方法もないわけではない。

[10] 旧バージョン 2.661 の入手方法は分かりにくいので、順を追って書いておく。Movable Type（http://www.movabletype.org/）から、your Account をクリックしてログイン（メールアドレスと本名でアカウントを無料で取得）すると、ダウンロードリスト（https://secure.sixapart.com/t/account）が表示されるので、2.661 を選んでダウンロードする（MovableType で行こう！ http://cheebow.sub.jp/docmt/）。なお対応する日本語化パッチはこちら（Milano::Monolog http://rebecca.ac/milano/mt/archives/000168.html）。

[11] 04 年 4 月、Movable Type（MT）の開発元である Six Apart（http://www.sixapart.jp/）は、日本のベンチャーキャピタル、株式会社ネオテニーからの出資を受け入れ、MT を利用した Type Pad サービス事業に乗り出した。これに伴い開発されたのが MT3.0 といってよい。個人利用は「ユーザが 1 名で 3 つまでの Weblog を作成する場合」だけに限定されていること、機能面での新味に乏しいこと等により、アップデートするユーザは多くなく、他の Weblog ツールへの移行が始まっている。このため、まもなく出る MT3.1 には何らかの機能を追加されるのではないかと期待されている。

[12] 画像添付メールを http://moblog.uva.ne.jp/ を経由して Weblog に投稿するシステム。特に携帯電話を経由する投稿スタイルが注目を集めている。

[13] Nucleus（ニュークリアス）3.1：http://japan.nucleuscms.org/ で入手可能。

[14] p_blog 1.0b20（安定β版）：http://pbx.homeunix.org/p_blog/index.php で入手可能。

[15] PukiWiki の UTF 化については本誌掲載の千田氏の論考で詳しく言及されているので、ご参照頂きたい。

[16] http://moli.cims.hokudai.ac.jp/chinavi/

[17] 図に見える「簡」「繁」というアイコンが前述のように <span lang="zh-cn"> といったタグを挿入するためのものである。二階堂善弘、千田大介諸氏の教示に基づき、Movable Type のテンプレートを改造した。

[18] 原文、訳文ともにコピーレフト。訳文はこちら。http://cruel.org/freeware/cathedral.html

[19]「伽藍とバザール」（同上）

特集2

N-gramによる漢字文献の分析

本誌の二号で「N-gram が開く世界　確率・統計的手法による新しいテキスト分析」という特集を組んだ。そこではシンプルな言語モデルである「N-gram」を利用することで、「これまでは見落とされてきたデータを見付けることができる」「従来“読み手の目”という視点で提示されてきた問題を数値的・統計的な手段で図示」するといった研究手法としての新たな可能性を、様々な方面から論じていた。

その後、本会代表の師茂樹氏が開発した UTF-8 対応の N-gram 共起頻度抽出ツール「morogram」により、従来ネックであった多漢字処理や人文系研究で見落とすことができない「頻度 1」集計が可能となった。まさしく「morogram」の登場は、漢字文献と N-gram モデルとの関係を新たな地平に誘うものとなったといえるだろう。そしてその後も、漢字文献を N-gram モデルにより分析した研究は着実に進められ、本誌にも関連する論攷がいくつも掲載されてきた。

本特集は、N-gram モデルを漢字文献でどう扱おうかと試行錯誤を始めたばかりの第二号の段階から、より深化した「N-gram を利用した漢字文献の分析」の現状を示すものである。更なる一歩を踏み出した N-gram による漢字文献分析の世界を存分に味わっていただきたい。

Contents

趙岐『孟子章句』の特徴

「注釈」と本文の語彙頻度比較を通した問題発見

齊藤　正高（さいとう　まさたか）

はじめに

小論の目的は『孟子章句』の「注釈」としての特徴を指摘することである。

方法としては、近年展開しつつあるコンピュータによる文字処理[1]をつかい、問題の解釈については、先行研究[2]から多くの示唆を得た。

なお、小論において、カッコ付きで「注釈」と書く場合は、古典のテキストに密着して、その解釈を行うこと一般を指している。

『孟子章句』の基本的性格

はじめに、テキストの名について若干の事項を、整理しておきたい。

まず、趙岐（109-201）[3]の『孟子章句』は、その執筆時点では、「注」ではなく、あくまで「章句」であった点を強調しておきたい。

書名の繰り返しにすぎない事実であるが、多くの書物が「趙注」[4]という呼び名で、『孟子章句』を論じており、厳密な用語を欠いたまま批評している。確かに、後世「注」という言葉は「解書之名」[5]を指す一般名詞として機能している。しかし、「注」が比較的新しい「注釈」のスタイルであった後漢にあっては、「章句」との差別化の意味をこめて、そう呼んでいたのである。だから、後漢末に執筆された『孟子章句』を問題にする場合、同時代に登場した、鄭玄（127-200）の「注」とは一応切り離して考えておかねばならない。そのうえで、「章句」と「注」を比較する基礎ができるのである。

そして、つぎに「章句」という名を正しく理解する必要がある。

揚雄[6]・桓譚[7]・班固[8]・王充[9]などが貶義として用いた「章句」は、前漢にさかえ、「師法を増して百萬言に至る」[10]と称された繁雑な「注釈」の伝統のことである。これは漢代の実学[11]と結びつき、その夾雑物が多くなり過ぎたために貶義をもつようになったと思われる。しかし、これは「章句」の本質ではない。

「章句」とは本来、分章・断句を基本[12]にして本文の解釈を行う師授の伝統である。そして、記憶の便のため、一部に有韻の句をもっていたと推定[13]されている。

『孟子章句』は5万字程度、本文に対して約1.4倍という分量であり、繁雑とすることはできない。したがって、『孟子章句』においては、まず純粋に解釈の手法として「章句」を理解しておけばいいのである。

この点、焦循は後漢末における「注釈」の形態に注意を払っており、『孟子正義』に以下のように指摘している。

> 毛詩傳全在矣、訓釋簡嚴、言不盡意、鄭氏箋之、則後世疏義之濫觴矣。鄭於三禮詳說之矣、乃周禮本杜子春鄭司農而討論、則

又後人集解之先聲也。何休公羊學、專以明例、故文辭廣博、不必爲本句而發。蓋經各有義、注各有體、趙氏於孟子既分其章、又依句敷衍而發明之、所謂章句也。

（『孟子正義』孟子題辭）

このように、「章句」と「注」が別の「注釈」であったことを理解し、「章句」の貶義をとりのぞくことができれば、『孟子章句』の内容のうち、「注」としては拙劣だと思われる点も、「章句」本来の機能として説明できるのである。例えば、以下のように、たいへん短く、かつ本文の繰り返しにすぎないかのような文の役割についてである。

『孟子』	『孟子章句』
曰有之	王曰有之（巻一）
曰否	王曰我不信也（同上）
許子冠乎	孟子問相（巻五）
曰冠	相曰冠也（同上）
謂王良良不可	王良不肖（巻六）
孟子曰否不然	否不是也（巻九）
曰然	告子曰然（巻十一）

表 1：『孟子章句』における短い「注釈」

これらは「章句」の中では、拙劣な「注」ではなく、断句のマーカーとして機能していると理解できる。そうすることで、問答体の多い『孟子』にこれらの文を挿入した意図を正しく評価できるのである。

また、次の例は異なる文に全く同じ文を添えている例である。

見孟子問曰周公何人也［賈問之也］曰古聖人也［孟子曰周公古之聖人也］（中略）曰周公知其將畔而使之與［賈問之也］曰不知也［孟子曰周公不知其將畔］（巻四、公孫丑章句下、燕人畔王章、［　］が趙岐）

これも「章句」の基礎作業である断句のマーカーと見なすことができよう。

また、本来『孟子章句』と一体であり、章末にあって章をまとめている『孟子章指』についても、分章をはっきり示している点を「章句」の機能として理解できる。なお、『孟子章指』の文に有韻の句がある点[14]も、師授にもとづく「章句」の本来の形態として理解できよう。

趙岐の生涯と学問

つぎに、趙岐の人生と学問形成ついて、分かっていることを見ておきたい。

趙岐は 109 年ごろ、京兆長陵（陝西省西安の北）の豪族の家に生まれ、初名は嘉であった。鄭玄や盧植の師である馬融（79-166）の姪[15]を娶ったが、外戚である馬融の士節をいやしんで往来は頻繁ではなかった[16]。

若いころ州郡に仕え、廉直で悪をにくんだため憚られた。三十歳ごろに大病をわずらい七年臥褥し、遺言を残したが快復した。

永興二年（154 年）司空であった房植の掾吏となった。房植は「黨人の議、此より始まる」[17]とされる清流の士人である。その後、大将軍梁冀に召され、献策をしたが容れられず、皮氏の長となり辺境を治め、学校を興した。だが、上役に宦官が赴任してくることになったので、職を辞して京兆に帰り、功曹に復職している。

延熹元年（158 年）、宦官系官僚の唐玹が京兆に赴任してくると、趙岐は従子の戩とともに逃亡することになる。彼は以前、唐玹を何度か貶議し、深く恨まれていたからである。唐玹は趙岐の家族を収監し、重罪に陥れ、尽く殺してしまう。趙岐は姓名を偽り、各地を放浪した。北海市中で餅を売っている時、孫嵩にみとめられ、彼の自宅の複壁に数年の間かくまわれた。唐氏が滅び、官界に復帰した後は、二度の党錮を生きぬき、黄巾の乱以後、敦煌太守・太僕などを歴任し、曹操の時、建安六年に九十歳余で卒した。

趙岐の師承は不明である。焦循は祖父に学んで常師はいないとの見解をとっている。ただし、『周官』を馬融から学んだ形跡があり、『孟子章句』にも本文と注を両載するなど、馬融の影響[18]を否定できない。焦循は「声音訓詁の学、馬（融）

鄭（玄）と殊ならず」[19] としている。

また、すでに弡和順氏の研究によって、趙岐が修めた詩経学が解明されており、『孟子章句』に『毛傳』と共通する例が 31 ヶ所指摘 [20] されている。さらに、『鄭箋』を参照していた可能性も極めて高いことが指摘 [21] されている。

鄭玄との共通点については、馬融を媒介とみなす見解 [22] がある。たしかに常師ではないにせよ、趙岐は馬融の影響をうけており、それが鄭玄との共通項であることも確かである。しかし、もうひとつの媒介もあったと思われる。趙岐が逃亡生活のなかで身を寄せていた孫嵩は、鄭玄と同郡であり、党錮のときも立場を同じくしており [23]、また鄭玄と同じ世代に属する。趙岐は若い死友 [24] 孫嵩を介して鄭玄の著作にふれたと考えることも、十分可能である。

『孟子章句』につけられた「孟子題辭」によれば、趙岐は複壁への避難を契機に『孟子章句』の執筆を始めたようである。後世、「患難顛沛」[25] の中で書かれたと理解される所以である。また、同じく「孟子題辭」に「敢えて達者に當たらざるも、新學に施し以て疑を寤り惑を辨ず可し」と述べ、いわば自らの「章句」を教科書的な基礎作業として評価している。

方法

以上のように、趙岐と『孟子章句』を理解した上で、その特徴を抽出する方法について述べたい。

小論で用いた手法は、「注釈」が依拠する本文をモノサシとしてあてることである。本文の語彙頻度と比較することで、注釈者がくり返し主張している点を、的確に取り出すことができ、そのなかから「注釈」としての特徴も浮かび上がってくるはずである。

本文と「注釈」の全面的比較には、コンピュータによる文字処理が有効である。そして、文字処理の出す結果が、たとえ意外なものであっても、その理由を追及せねばならない。そうすることで、今まで知られなかった傾向も見えてくるであろうし、読解の中で捨象されてきた事柄もはっきり示すことができるのである。

コンピュータによる処理には、データの構築と数量化という二つの面がある。以下に順に述べる。

■ データの構築

底本は、四部叢刊所収「清内府蔵宋刊本」[26] を用い、分析用テキストデータを作った。四部叢刊本は阮元本より通じやすい点が多く、阮元本では校勘記に出ている『孟子章指』が一つ一つの章末に付いており、趙岐が書いたテキストにより近いと思われるからである。

グリフを正確に再現できなかった文字や、検索用に統一した文字には以下がある。

データ	底本の文字
走	上部が大、ソウニョウも同じ
弘	欠画あり
完	欠画あり
殷	欠画あり
匡	欠画あり
肯	止の横画が冖の字を統一
汚	汙・汚を統一
姦	姦の右に干がある字を統一
胸	胷と[illegible]の字を統一
他	「佗日」一例を「他日」に統一

表 2：分析用テキストデータにおける文字統一例

最終的には以下のようにマークアップし [27]、XML 文書を作成した。

```
<?xml version="1.0" encoding="UTF-8"?>
<mengzi><volume>
（中略）
<text no='215' vol='13'><ph><scr>孟子曰形色天性也</scr><com>形謂君子體貌嚴尊也尚書洪範一曰貌色謂婦人妖麗之容詩云顏如蕣華此皆天假施於人也</com></ph><ph><scr>惟聖人然後可以踐形</scr><com>踐履居之也易曰黃中通理聖人內
```

```
外文明然後能以正道履居此美形不言居色主名
尊陽抑陰之義也</com></ph><sum>章指
言體德正容大人所履有表無裏謂之柚梓是以聖
人乃堪踐形也</sum></text>
（中略）
</volume></mengzi>
```

<mengzi>はXML文書のルート要素、<volume>は巻をまとめているタグである。<text>は章をあらわし、<ph>は章のなかの句をあらわす。<scr>は『孟子』の本文、<com>は趙岐の「注釈」、<sum>は『孟子章指』を表す。

このような構造化は便宜的なものであるが、作成したXML文書にXSLT[28]を施せば、ブラウザーでそれぞれの要素を任意に取り出すことができる。また、JavaScriptなどでDOM[29]を操作するスクリプトを書くことでも、同様の操作を行うことができる。

```
例：
shousi.xsl （章指のみを抽出するXSL）
<?xml version="1.0"
encoding="UTF-8" ?>
<xsl:stylesheet version="1.0"
xmlns:xsl=
"http://www.w3.org/1999/XSL/
Transform">
<xsl:template match="/mengzi">
<xsl:for-each select="volume">
<xsl:for-each select="text">
<xsl:value-of select="sum"/><br/>
</xsl:for-each></xsl:for-each>
</xsl:template></xsl:stylesheet>
```

分析用テキストデータの基礎統計は以下になる。

	『孟子』[30]	『孟子章句』	『孟子章指』
文字数	35,451	50,557	8,298
文字種	1,904	2,344	1,188

表３：分析用テキスト各部分の基礎統計

注記：

巻頭の「孟子題辭」および巻末の「孟子篇叙」は比較すべき本文がないので除外した。また、『孟子』については篇題の文字も文字数に含めた。篇題の下に「章句」が付けられているからである。

■　数量化

特徴の抽出には、以下の比を用いた。この値をかりにCSR（Commentary-Scripture Ratio）としておく。

```
CSR=CF／SF
CF:注釈における文字頻度（注釈頻度）
SF:本文における文字頻度（本文頻度）
```

CSRは単純に「注釈」において或る文字が本文の何倍つかわれているかを示す値である。これだけでも十分、「注釈」に独特な文字を抽出することが可能だが、「注釈」と本文の両者に濃度（Concentration）の観点を導入し、濃度比CSRCを求めると、次の式になる。

```
CSRC=(CF／SF)・(SL／CL)
CF:注釈頻度　SF:本文頻度
SL:本文長　　CL:注釈長
```

分析用テキストデータにおいて「仁」のCSR及びCSRCは、以下である。

```
CSR仁=211／158=1.34
CSRC仁=(211／158)・(35451／50557)=0.94
（少数点以下第3位を四捨五入）
```

単純に頻度だけでは、「仁」は本文より、およそ1.34倍使われていることになる。だが、濃度を考慮すると、比は0.94になり、『孟子』における「仁」より淡いことを示している。つまり、『孟子』には趙岐よりも熱心に「仁」を言う傾向があるのである。

このCSRとCSRCには使用範囲がある。本文

と「注釈」双方に使われている文字にしか使用できないという点である。もし、本文頻度や注釈頻度が0であれば、CSRやCSRCの値も、0や計算不能になって比較を行うことはできなくなる。

したがって、0頻度の文字については別に考察を行う必要がある。

結果

以下、分析用テキストデータを数量化して特徴を抽出した結果を示す。結果そのものが繁多であるため、あらかじめ、その一部をとりあげて検討することを断っておきたい。

■ 1-gramからみた特徴

『孟子章句』に『孟子』より10倍以上使われている文字をCSRの大きいものから順に並べると以下になる。

	『孟子』	『孟子章句』	CSR
意	2	71	35.50
各	1	25	25.00
遭	1	23	23.00
須	1	17	17.00
魏	1	17	17.00
逸	1	16	16.00
賦	2	31	15.50
困	1	15	15.00
喩	3	44	14.67
履	1	14	14.00
邑	2	23	11.50
論	2	23	11.50
名	10	111	11.10
勉	1	11	11.00
量	1	11	11.00
迫	1	10	10.00
勸	2	20	10.00
譬	2	20	10.00
驕	1	10	10.00

表4：CSRの高い文字

用途が簡単に分かる文字については先に述べておきたい。

- 引用の書名に含まれている文字
  → 逸・論（『尚書逸篇』・『論語』など）
- 地名を示すために使われている文字
  → 魏・邑(「周霄魏人」・「平陸齊下邑也」など)

つぎに、最も比が大きく本文に対して35倍も使用されている「意」の用例と、『孟子章句』において多用され、かつ本文との比率も高い「名」の用例を検討する。

■ 意

「意」の用例は、多岐に渡るが、ある人物の意図として使われている例が多数あり、これを『孟子章句』における一定の傾向として指摘できる。

以下に分類を示す。なお、引用に付けた番号は趙岐の分章における『孟子』の章を指し、全261章である。

- 「王意」5例
  - 3：因王好戰故以戰事喩解王意
  - 7：孟子雖心知王意而故問者欲令王自道緣以陳之
  - 7：王意大而不敢正言
  - 7：言王意欲庶幾王者莅臨中國而安四夷者也
  - 140：王意解顏色定復問異姓之卿如之何
- 「孟子之意」「孟子意」4例
  - 23：孟子之意以爲魯侯欲行天使之矣
  - 41：欲爲王解孟子意
  - 42：孟子意欲使王繼
  - 42：季孫知孟子意不欲而心欲
- 「其意」9例（意図と意味の例が混在している）
  - 10：王謂孟子之言大不合於其意答之云寡人有疾
  - 25：丑曰不動心之勇其意豈可得聞與
  - 55：孟子言人但破碎瓦畫地則復墁滅之此無

用之爲也然而其意反欲求食則可食乎
58： 言其意苦勞極甚於仲夏之月治畦灌園之勤也
91： 每人輒欲自加恩以悅其意則日力不足以足之也
123：因以萬章具陳其意耳
165：屋廬子已曉其意聞義而服故悅也
188：殺此罪人者其意欲生民也
216：因公孫丑使自以其意問孟子

●「告子之意」「晏子之意」
146：或人以爲可教以善不善亦由告子之意也
11： 晏子之意不欲使景公空遊於琅邪而無益於民也

●「聖人之意」（孔子を指す）
260：孟子錄之以其不解於聖人之意故謂之萬子

これらは、趙岐が『孟子』を解釈する際に、本文にあらわれる人物の意志を忖度してつけた「注釈」である。このような用法は実は『孟子』に根拠がある。『孟子』には「意」が２例つかわれている。

86： 孟子謂、樂正子曰、子之從於子敖來、徒餔啜也、我不意、子學古之道而以餔啜也（巻七離婁章句上）
126：故說詩者、不以文害辭、不以辭害志、以意逆志、是爲得之、如以辭而已矣、雲漢之詩曰、周餘黎民、靡有孑遺、信斯言也、是周無遺民也（巻九萬章章句上）

126の下線部が、「意」を章句に多用する理由である。趙岐は「孟子題辭」にもこの句を用いている。

其言曰說詩者、不以文害辭、不以辭害志、以意逆志、爲得之矣、斯言殆欲使後人深求其意、以解其文、不但施於說詩也。

「不但施於說詩也」と述べている通り、趙岐は自らの『孟子』解釈においても、「以意逆志」の解釈法を応用したものと考えられる。

つぎに「天意」という用例を指摘しておきたい。

●「天意」５例
23： 行止天意非人所能爲也
40： 天吏天所使謂王者得天意者也
62： 言君無道術可以揆度天意
127：當與天意合之非天命者天子不能違天命也
261：然而世謂之無有此乃天不欲使我行道也故重言之知天意之審也

40（巻四、沈同以其私問章）の例は、王者と「天意」の関係を述べたもので、『春秋繁露』深察名號篇に「受命之君、天意之所予也」、「王承天意以成民之性為任者也」とあり、影響をうかがわせる。

『孟子』には、「天意」は使われていない。また、経書にもない言葉である。集中的に使われているのは、『春秋繁露』のほかに、『墨子』天志篇、『大平經』、『尚書僞孔傳』などである。後漢の「注釈」では、鄭玄『毛詩鄭箋』6、『周禮注』1、『禮記注』1、何休『春秋公羊傳解詁』4である[31]。「天意」という言葉の成立のなかで、『墨子』と漢代の思想がどのような関係にあるのか興味ぶかい点であるが、これについては別の機会にしたい。

思想的に見ると、趙岐は40・62・127に「天意」を為政者の権力を根拠づけるものとしている。23・261の「天意」は孟子自身の願望が成就しないことに関係している。後者の「天意」は、運命といいかえることも可能で、ひろく人間に「禍福」をもたらすものであろう。この点は本田濟氏が趙岐の思想として挙げる「徳には福が伴って然るべきだという信念」[32]とも関わってくるであろう。

■ 名

「名」の使用例を見てみると、人名29・地名7・篇名5・水名3・木名3・国名2・山名1が挙げられ、全体の約半数をしめる。これらは「注釈」に必要

な説明であり、これによって頻度が高くなっているのも納得できる。

残りの用例のなかで、開巻第一の梁惠王問利章にふくまれる問題を指摘しておきたい。

『孟子』の本文は以下である。

孟子對曰、王何必曰利、亦有仁義而已矣

これに対して趙岐は以下のように解釈している。

1：　孟子知王欲以富國強兵爲利、故曰王何必以利爲名乎、亦惟有仁義之道者、可以爲名、以利爲名、則有不利之患矣、因爲王陳之。

まず、本文にはない「名」を導入している点に注目すべきである。ここによく似た例に巻十二、宋牼章の『孟子章句』がある。

164：以仁義之道不忍興兵、三軍之士悦、國人化之、咸以仁義相接、可以致王、何必以利爲名也。

これらの例を「孟子は決して根本的に利に反対したのではなく、利を名号とすることに反対したのだ」[33]と、利の擁護に読む理解もある。しかし、1を「治國之道、明當以仁義爲名」と要約している『孟子章指』をみれば、重点はそうでないことが分かる。

朱熹は1と同じ本文を以下のように解釈する。

仁者、心之德、愛之理、義者、心之制、事之宜也。

比較を通してみると両者の差異は判然とする。つまり、朱熹は「仁義」を「心」を媒介にして人間性として理解しているが、趙岐は「仁義」を「仁義之道」と考え、「名」を媒介にして政治理念として理解していること分かるのである。

趙岐は後述のように「仁義之道」を多用する。この政治理念を『孟子』のなかに求めた背景には、濁流との闘争のなかで家族を失い、自身の「名」をも失った趙嘉の前半生がある。そして、この理念は晩年、崩壊していく漢帝国の末期にあって、隠逸という方法[34]をとらず、まさに老骨に鞭うって、政治の場にありつづけた趙岐の行動を裏づけるものであったと思われる。

■　注釈頻度が0の文字について

つぎに、趙岐が使わなかった『孟子』のなかの文字をみてみる。『孟子』における頻度3以上のものを挙げると以下になる。

惻　煖　遍　俯

この中で「惻隱之心」として、『孟子』に5回出現する部分に、趙岐が「章句」をつけていない点が問題となる。以下は、はじめて「惻隱之心」がでてくる例である。

『孟子』

今人乍見孺子將入於井、皆有怵惕惻隱之心非所以内交於孺子之父母也。（公孫丑上）

『孟子章句』

乍暫也、孺子未有知小子、所以言人皆有是心、凡人暫見小小孺子將入井、賢愚皆有驚駭之情、情發於中、非為其人也、非惡有不仁之聲名、故怵惕也。

趙岐『孟子』を「拙にして不明」[35]と評価した朱熹の『孟子集注』には、この部分に「惻傷之切、隱痛之深也」とある。焦循『孟子正義』には『説文解字』の「惻痛也」をあげ、「惻隱皆痛也」（『漢書』鮑宣伝の顔師古注）をあげる。

趙岐は『毛傳』にもとづいて「隱痛也」（梁惠王上）としている部分があるから、焦循が顔師古注によって示した訓詁の一部は存在しているのである。

しかし、許愼（30～124）の『説文解字』の説はとっていない。そこで、焦循によって既に共通点が指摘されている鄭玄の「注」を参照してみ

ると、以下の例があることに気づく。

惻隱之心能如是、則其餘齊衰下直道而行盡自得也。（『禮記』雜記「色喪之外」鄭玄注）

ここの経文には「惻隱」はないから、「惻隱」が説明を要する言葉であれば、このようには「注」をつけないだろう。また、鄭玄が引く鄭衆（？～89）の説には以下がある。

鄭司農云、測讀為惻隱之惻。
（『周禮』冬官弓人「漆欲測」注）

このような声音の説明は「惻隱」が人口に膾炙していなくては成立しない。したがって、鄭衆や鄭玄には「惻隱」は説明不要な言葉だったのである。これが趙岐にも共通する感覚であるが、趙岐が許愼の説をとらないのは、鄭玄が『駮五經異義』を書いて許愼に反論したという当時の風潮もあるのかもしれない。

■　本文頻度が 0 の文字について

これは趙岐だけが使っている文字であるから、『孟子章句』の解釈方法を直接示しているものが多い。以下に頻度の高いものから代表的な用途を示す。

●頻度 20 以上

篇　但　衆　雅　肯　怪

● **「篇」「雅」**
→詩の篇名を挙げる場合に使用している。

● **「但」**
→「先秦以前に比較的用例が少ない」[36] と指摘がある。

● **「衆」**
→「人」や「國人」を「衆」や「衆民」「衆人」といいかえる例が多い。

● **「肯」**
→大多数が「不肯」と使われており、趙岐が本文にでてくる人物の意図について強調をしている例が多い。

● **「怪」**
→「王怪其大」のように、王や孟子自身、孟子の弟子が問いを発した部分の意図を説明している例が多い。

●頻度 19 ～ 10

隨　滿　恨　貪　念　記　循　謚

● **「恨」**
→「憾恨也」という訓詁が 2 例ある。この訓詁は『毛傳』には存在しないが [37]、鄭玄の『禮記』注 [38] には存在する。「孟子去齊有恨心」（公孫丑下）、「蓋恨其不得學於大聖也」（離婁下）など、孟子の言動からその心情を推しはかっている部分に多く用いられている。

● **「貪」**
→「頑貪」という訓詁がある。これは趙岐独自の訓詁である。『章指』においても「貪」は 4 例を数え、本文の「不嗜殺人」を「不貪殺人」に書き換えた例がある。

● **「記」**
→「禮記」や「魯史記」などの用例のほかに、「志記也」の訓詁がある。これは鄭玄の『周禮』注にみえる訓詁 [39] である。

● **「循」**
→「率循也」という訓詁があり、すでに先行研究によって『鄭箋』との類似が指摘 [40] されている。また「遵循也」[41] という訓詁もあり、これは『毛傳』に従っている。

● **「謚」**
→「襄謚也」「宣謚也」「平謚也」など王公の謚号を示す字である。

以上、用途を限定できない字もあるが、用途を示すことができた例では、『毛傳』や鄭玄の学問との共通点が認められる。『鄭箋』に限らず、『周禮注』や『禮記注』と同様の訓詁を『孟子章句』が保有していることは、焦循の指摘を確認するものである。

また、「怪」「恨」など『孟子』にはない文字を援用して、『孟子』に書かれた人物の心情を忖度して「注釈」をほどこす手法は「以意逆志」の延長線上にあるものであろう。

■ N-gramからみた特徴

最後に、N-gramツール[42]をつかい、数値の大きいgram数で共起頻度を集計した結果を示す。これによって、趙岐が用いる頻度の高い句が分かる。

孟子弟子	17
仁義之道	15
君子之道	14
尚書逸篇	9
聖人之道	9
堯舜之道	6
學於孟子	6
仁義禮智	5
天下之民	5
生民以來	5
仁義之心	4
今之諸侯	4
以利爲名	4
武王伐紂	4
簞食壺漿	4

表5：『孟子章句』の高頻度句（右は頻度）

一見して分かる傾向は「仁義之道」「君子之道」「聖人之道」「堯舜之道」など「之道」をふくむ句の頻度が高いという事実である。整理すると以下になる。

表6：○○之道

句	CF	SF	CSR
仁義之道	15	0	
君子之道	14	2	7.00
聖人之道	9	1	9.00
堯舜之道	6	9	0.67

「仁義之道」は前述のとおり「名」としてあらわれる政治理念である。これは『孟子』にはない言葉である。経書では『禮記』禮器にある。趙岐はほかにも「曲禮」や「王制」から5例引いているので、趙岐の禮に対する考え方と政治理念との関係を考える材料となるだろう。また、「堯舜之道」が『孟子』のような熱心さでくり返されていない点も分かる。

つぎに、『孟子章指』についてもみてみる。

賢者志其大者	3
人君恭険	2
君子之道	2
君子小人	2
大人之行	2
時行則行	2
爲仁由己	2
移風易俗	2
聖人之道	2
聖人量時	2
舎生取義	2
行仁天下	2
見幾而作	2
賤不失道	2
金石獨止	2

表7：『孟子章指』の句（右は頻度）

『章指』は章のまとめであるから、繰り返し表現自体が少ない。また、全体に通用する表現は少なく、それぞれが章の独自性をもっている。その中で、くり返し表れている句は『孟子章指』の外部に典拠がある場合が多い。この点は、焦循『孟子正義』によってほぼ解明されている。

「賢者志其大者、不賢者志其小者」
　（『論語』子張、「賢者識其大者」）
「時行則行、時止則止」「時行則行、時舎則舎」
　（『周易』艮、彖伝「時止則止、時行則行」）
「爲仁由己」
　（『論語』顔淵）
「移風易俗」
　（『禮記』樂記および『孝経』）

「舍生取義」
（『孟子』告子上「舍生而取義者也」）
「君子見幾而作」「見幾而作、不俟終日」
（『周易』繋辭下）
「萬物皆流、而金石獨止」
（『説苑』説叢、「水浮萬物、玉石留止」）

まとめ

以上、計量的観点から『孟子章句』の特徴の一端を示した。

趙岐による『孟子』の解釈法とかかわる本文比最大の「意」の問題、その少数例が趙岐の政治理念とつながる「名」の問題、趙岐の言語感覚にかかわる0頻度の「惻隠」の問題などは、みな数量化を経て、はじめて問題として明確に指摘できるものである。

また、先行研究によって指摘されてきた鄭玄との共通点も、新たな根拠をもとに確認することができた。

小論の手法は、「注釈」と本文の頻度比に注目するという非常に単純なものである。だが、数量化したテキストの情報を問題発見の手がかりとして使うことで、条件つきであっても、事実としてのテキストの姿をとらえることが可能になると思われる。

今後、このような計量的なテキストの分析をすすめていくには、ある程度、XMLによって構造化されたテキストの作成が重要になるであろう。

注

[1] 『漢字文献情報処理研究』第2号のNグラム関連論文全て、拙稿「TFIDFの漢文テキストへの応用」（第3号）、漢文における本格的なNグラム分析の成果である秋山陽一郎氏「『老子』傅奕本来源考」（第4号）、また、新たな領域を開拓している文字オントロジーの基礎研究CHISEプロジェクトなどの報告を参照されたい。

[2] 趙岐と『孟子章句』に関する先行研究は管見の及ぶ所では以下の通り。
本田濟「趙岐『孟子章句』について」1980（『池田末利博士古稀記念東洋学論文集』所収）
䒦和順「趙岐とその学問」1985（北海道大学中国哲学会『中国哲学』第14号所収）、
同前「趙岐の詩經學」1996（詩經學会『詩經研究』第20号）
董洪利『孟子研究』江蘇古籍出版社 1997
杉山一也「趙岐撰『孟子章指』の輯佚」1988（『中国研究集刊』宙号所収）

[3] 前掲、本田濟1980の説をとった。

[4] 『隋書』經籍志にも「趙岐注」とある。また以下の例がある。
「趙注尚異」（偽孫奭疏）
「趙岐之學、以較馬鄭許服諸儒、稍為固陋。然屬書離辭、指事類情、於詁訓無所戾、七篇之微言大義、藉是可推。且章別為指、令學者可分章、尋求於漢傳注、別開一例、功亦勤矣。」（阮元『孟子注疏校勘記序』）
「是注即岐非難北海時、在孫賓家、夾柱中所作。漢儒注經多明訓詁名物、惟此注箋釋文句、乃似後世之口義、與古學稍殊。」（『四庫全書總目提要』孟子正義）

[5] 「注者即解書之名、但釋義之人多稱為傳、傳謂傳述為義、或親承聖旨、或師儒相傳故云傳、今謂之注者謙也、不敢傳授、直注己意而已、若然則傳之與注各出己情、皇氏以為自漢以前為傳、自漢以後為注、然王肅在鄭之後、何以亦謂之傳、其義非也」（『禮記正義』巻一、曲禮疏）

[6] 「章句之徒相與坐而守之」（『漢書』揚雄傳）

[7] 「徧習五經、皆訓詁大義、不為章句」（『後漢書』桓譚傳）

[8] 「所學無常師、不為章句、舉大義而已」」（『後漢書』班固傳）

[9] 『論衡』書解篇

[10]『漢書』儒林傳

[11] 野村茂夫「前漢『章句の学』試探」1978、愛知教育大学研究報告27所収

[12]「離章辨句、委曲枝派也」（『後漢書』桓譚傳、李賢注）

[13] 小南一郎「王逸『楚辭章句』をめぐって一漢代章句の學の一側面」1991、『東方學報』第六十三冊所収

[14] 前掲、杉山一也氏1988

[15]『後漢書』趙岐伝李賢注所引『三輔決疑注』にしたがって、従姉妹とする説もある。䒦和順氏1996 p.7

[16]『後漢書』趙岐傳

[17]『後漢書』党錮列傳

[18] 焦循『孟子正義』孟子題辭には、馬融に対して「岐雖

鄙融之爲人、而義有不通、亦往請問、則其虚心取善可知」とする。

[19] 焦循『孟子正義』孟子題辭

[20] 前掲、弛和順氏 1985

[21] 前掲、弛和順氏 1996

[22] 前掲、弛和順氏 1996

[23]『後漢書』鄭玄傳

[24]『後漢書』趙岐傳

[25]「黄覇之受尚書、趙岐之注孟子、皆在患難顛沛中」（王應麟『困學紀聞』卷一）

[26] 巻二下に「至正二十五年二月」（1365 年）の書込あり。

[27] http://taweb.aichi-u.ac.jp/saitom にて公開し、諸賢の批正を請う予定である。

[28] eXtensible Stylesheet Language Transformations の略

[29] Document Object Model の略

[30]『孟子』の字数については、「孟子題辭」34685 字、陳子元「孟子雜記」35411 字、周廣業「孟子異本考」35226 字（いずれも『孟子正義』所引）

[31] http://www.sinica.edu.tw/ftms-bin/ftmsw3、台湾中央研究院漢籍電子文獻を使用した。

[32] 前掲、本田濟氏 1980

[33] 前掲、董洪利氏 1997、p.167-168

[34] 隠逸君子の郷村社会における役割については、川勝義雄「漢末のレジスタンス運動」1967『東洋史研究』24-5 所収を参照した。

[35]『朱子語類』巻五十一

[36] 中国社会科学院語言研究所古代漢語研究室編『古代漢語虚詞辞典』商務印書館 p.84

[37]「悔恨也」という訓詁が大雅「雲漢」の毛伝にある。

[38]『禮記』中庸、「人猶有所憾」注

[39]『周禮』春官、外史「掌四方之志」注

[40] 前掲、弛和順氏 1996

[41] 梁惠王下、「齊宣王見孟子於雪宮」章

[42] 師茂樹氏 morogram（http://sourceforge.jp/projects/morogram/）を使用した。

中国戦国期の語彙量について

N-gram とユールの *K* 特性値を利用した分析

山田　崇仁（やまだ　たかひと）

はじめに

中国の戦国期（前 403 ～前 221）は、政治的には春秋期以来の周王→覇者→諸侯という政治体制から、各国称王に伴う専制国家へと変化し、ついには天下統一に至るという中国史上有数の激変の時代であった。

また思想的な面でも中国史上最も多様な思想が説かれた時代で、これを『史記』屈原賈生列伝では「諸子百家」と称しているが、「諸子百家の言」を携える「遊説の子（遊子）」は、自らが提唱する新たなる概念を表現するために造字・造語を盛んに行い、それを武器として各国を巡り自説を展開していったのである。

その世界では、「共通に用いる言葉」「流行と衰退とが僅かな間で変化した言葉」「特定の学派・時代・地域のみで使われた言葉」といったさまざまな言葉が飛び交い、まさに「言葉の戦場」といったおもむきすらあった。

従来、諸子百家の用いた言葉については、個々の思想的特徴や各諸子書（諸篇）の成書年代を想定するという観点で注目されてきた。

それは個々の言葉に関する分析は精緻を極めつつ、「儒家的」「道家的」といった個別の枠組み、あるいはカールグレンのいわゆる「前三世紀の標準文語」といった形での枠組みとして提示されてきたが、それぞれの諸子百家がどのくらいの言葉を保持（使用）していたかについて、定量的な形で諸子百家の言葉を把握しようとした研究は見られなかった。

一体、諸子百家の本質は「言葉を媒介として自らの主張を世に広める」にある。従って、彼らの保持していた言葉を定量的な形で把握することは質それぞれの学派あるいは時代を把握する上で必要なことである。

「諸子書の使用した言葉の量」という観点から諸子書を比較すると、例えば『孟子』（前 4 世紀末）や『荘子』逍遥遊・斉物論（前 4 世紀末～前 3 世紀初頭）は、『論語』（前 5 世紀末）や『礼記』の坊記・緇衣（前 4 世紀前半）といったそれ以前に成書された書物と比較した場合、言葉の種類や文章の長さにおいて後者の方がより素朴に感ずる。特に『荘子』の言葉は、儒家とは別世界に属しているといえるだろう。

更に戦国末期の前 3 世紀半ば～後半になると、例えば『荀子』では「人主」「名主」といったそれ以前の儒家では使用されなかった言葉が使われる一方、本来道家系の言葉である「萬物」を頻用するといったように [1]、特定の諸子で使われた言葉を他の諸子でも使用することで、言葉（ここではあくまで書き言葉（文語）を意味する）が一定の共通化の方向性を見せ始めるといえる。

但し、これはあくまでざっと読んだ上での印象

的な記述にしか過ぎない。言葉（文字・語彙）の変化を時間・学派といった観点で定量的に把握するためには、これらを何らかの形で数値化して比較すべきだろう。そこで本論では、N-gramモデルを用いて共起と共起頻度を集計した結果をユールの*K*特性値（語彙量を量る統計指標）を使用して数値化し、それを比較することにした。

本論の分析手法について

■ *K*特性値とは？

*K*特性値（*K-Characteristic*）とは、統計学者のユール（George Udny Yule,1871-1951）によって提唱された語彙量を表す統計指標である。

*K*特性値は、文書中にx_i回現われた語の個数をf_iで表すとき、以下の式で導き出される。*K*特性値を求める際に10の4乗（10,000）するのは、*K*特性値の値が小さくなりすぎるのを防ぐためである。

*K*特性値の算出式は以下の通りである。

$$K = 10^4\left(\frac{S_2 - S_1}{S_1^2}\right)$$

ただし、S_1とS_2との値は以下の式により求める。

$$S_1 = \sum_{x=1}^{x\text{の最大値}} (i \times f_i)$$

$$S_2 = \sum_{x=1}^{x\text{の最大値}} (i^2 \times f_i)$$

例えば1,000の単語で構成されるテキストを対象とした場合、単語の種類が全て異なっているとすると*K*特性値は0となり、全て同一だとすると9,990となる。このように*K*特性値は語彙量が多いほど数値が小さくなり、少ないほど数値が大きくなるという特徴を持っているため、語彙量の豊富さあるいは語彙の集中度（パターン化）の度合いを調査する際に用いられている。

通常、*K*特性値はテキストの単語あるいは形態素を対象として求めるが、本論では漢字が原則として一文字毎に言葉の概念を持つ「表語文字」[2]であるという特徴と、古典漢文がそもそも分章や単語の区切りすら議論の対象となるという点を踏まえ、N-gramモデルを利用して共起頻度を集計し、*K*特性値を求めることにした。

■ N-gramとは？

N-gramモデルは、情報理論の創始者として名高いクロード・エルウッド・シャノン（Claude Elwood Shannon）が提唱した確率統計的な言語処理の手法である。

N-gram自体は「あるテキストの総体を前から順に任意のN個の文字列または単語の組み合わせで分割したもの」を意味する。N個を単位とする文字列または単語の組み合わせを「共起」と呼び、N個の数（gram数）に応じてそれぞれ「1-gram,2-gram,3-gram…」と呼ぶ（二文字単位ならば2-gramである）。またテキスト全体での任意の「共起」を集計した結果は「共起頻度」と呼ぶ。

上述のように、漢字は「表語文字」である関係上、基本的には一文字が形態素の単位で分けられているという特徴を持つ。そのため、他の言語なら余り意味のない1-gramでも有意な結果を得る事が出来る反面、大きい単位でのN-gramを対象とする場合、ノイズと呼ばれる無意味な共起用例が格段に増加するため、結果として共起頻度が少ないデータしか得られず、複数文献を対象とした共起頻度の比較をする場合において、有意な結果を得られない可能性が高くなる。

同様にN-gramを利用して*K*特性値を求めた場合でも、gram数が大きくなればなるほど、単語や形態素単位での*K*特性値と比べて数値が極端に小さくなる。これは、N-gramが語彙という「意味のある単位」ではなく「のべ組み合わせ（意味のない単位も含む）」の集計結果に基づいて*K*特性値を求めるためであるが、それでもやはり文献がより少ない（多い）文字または語彙の組み合わせで構成されている（パターン化の度合いが高い）場合は、*K*特性値が意味のある結果を導き出すは

ずである。そのため本論では、1 ～ 5-gram までを対象として *K* 特性値を求め、文献毎の特徴はもとより、gram数別の傾向も分析することにした。

一般的に、*K* 特性値は対象となるテキストの長さが長ければ長いほど値が小さくなる傾向があり、計算の際には分析対象とするテキストの長さを統一することが望ましいとされる。

そこで本論では各テキストの「先頭から 1000 字＋意味の切れる部分」までを対象とし、その上で 1 ～ 5-gram で共起頻度を集計し、そこから *K* 特性値を求めている。なお、テキストの新字・旧字などの異体字についてはどちらかに統一し、句読点は無視している。

■　対象となるテキスト

本論で分析対象として用いるテキストは以下の通りである。

- 『論語』学而＋為政・述而・顔淵
- 『礼記』緇衣・坊記・表記
- 『孟子』梁恵王・公孫丑・万章
- 『荀子』修身・王制・王覇・正名・非十二子・礼論・堯問
- 『墨子』尚同上・尚同中・尚同下・兼愛上・兼愛中・兼愛下・天志上・天志中・天志下
- 『荘子』逍遙遊・斉物論・天地・天運・庚桑楚・天下
- 『呂氏春秋』有始覧・孝行覧・慎大覧
- 『韓非子』姦劫弑臣・顕学・孤憤・説難

これらは、儒家・道家・墨家・法家・雑家という学派的基準、及び前 5 世紀末～前 3 世紀後半という戦国期の諸子文献の成書年代の幅という基準で選択したものである。以下に、学派・時代別分類を挙げた。

●学派別

儒家
　『論語』『礼記』『孟子』『荀子』
墨家
　『墨子』
道家
　『荘子』
雑家
　『呂氏春秋』
法家
　『韓非子』

●時代別

前 5 世紀後半
　『論語』[3]
前 4 世紀前半
　『礼記』緇衣・坊記・表記 [4]
前 4 世紀後半
　『孟子』[5]『荘子』逍遙遊・斉物論 [6]
前 3 世紀後半
　『荘子』天地・天運・庚桑楚・天下
　『荀子』[7]『韓非子』[8]『呂氏春秋』[9]
前 3 世紀
　『墨子』[10]

N-gram の共起頻度算出には、花園大学の師茂樹氏の作成・極悪氏改造のコンピュータプログラム「morogram」を、*K* 特性値の算出にはこれも師氏作成の perl script スクリプト k-characteristics.pl をそれぞれ使用した。[11]

■　作業の一例

では次に、「『論語』学而＋為政」を例として N-gram ＋ *K* 特性値を求める作業を説明しよう。

以下に挙げたのが、分析対象となるテキストの一部である。

テキストは上述のように異体字や句読点を加工してあるが、句読点を省略したのは、1000 字という文字数に重きを置いたのが理由である。従って句読点込みのテキストを分析した場合に比べ、*K* 特性値の値は小さくなる（特に 4 ～ 5-gram ではその傾向が大きい）。

子曰學而時習之不亦説乎有朋自遠方來不亦樂乎人不知而不慍不亦君子乎有子曰其爲人也孝弟而好犯上者鮮矣不好犯上而好作亂

者未之有也君子務本本立而道生孝弟也者其爲仁之本與子曰巧言令色鮮矣仁曾子曰吾日三省吾身爲人謀而不忠乎與朋友交而不信乎傳不習乎子曰道千乘之國敬事而信節用而愛人使民以時子曰弟子入則孝出則弟謹而信汎愛衆而親仁行有餘力

このテキストを対象に、「morogram」を利用して「2-gram（二文字単位）」「共起頻度1以上」で集計する。

表1に、morogramの出力結果から、共起頻度上位20位までを挙げた。左から「共起頻度・文字列・gram数」である（例えば「子曰」は38回出現するということを意味する）。

次にmorogramの出力結果から、k-characteristics.plを利用して*K*特性値を導き出す。

表2が実際の出力結果である。それぞれの記号・数値の意味を解説しよう

表1：2-gram 異なり字数と延べ字数の比較

38	子曰	2
8	之以	2
8	而不	2
7	君子	2
5	不知	2
5	以禮	2
5	十而	2
5	問孝	2
5	子貢	2
5	矣子	2
4	其所	2
4	孝子	2
4	曰君	2
4	爲政	2
4	而無	2
4	與子	2
4	貢曰	2
3	不亦	2
3	之不	2
3	之與	2

x	f
1	700
2	69
3	15
4	7
5	6
7	1
8	2
38	1

S1 = 1002
S2 = 2994
K = 19.8405584041498

表2：k-characteristics.plの出力結果

まず、xは共起頻度、fは各共起頻度の文字種数を示す。

表1を見ると、xが「38」となるのは「子曰」のみだからfは「1」、同様にxが「8」となるのは「之以」「而不」の二種類だから、fは「2」となる。

S_1は上述の計算式を使うと（1 × 700）＋（2 × 69）…（38 × 1）となり、値は「1002」である。同様にS_2も（1^2 × 700）＋（2^2 × 69）…（38^2 × 1）で、値は「2994」となる。従って*K*特性値は「19.841（小数点4桁目で四捨五入）」となる。

$$K = 10^4\left(\frac{2994 - 1002}{1002^2}\right) = 19.841$$

次ページの表3が、上記の方法で*K*特性値を算出したものの一覧である。更に参考用として、「個別の文字種の字数」と「延べ字数」を比較したものも挙げた（表4）。こちらは、各諸子書から12,600字前後（延べ字数）を抽出し、それを1-gram単位で共起頻度を集計し、次にこれを「異なり文字種と文字数」「全ての諸子書で共用される文字種と文字数」「個別の諸子書にのみ使用される文字種と文字数」に絞り込んだものである。

テキスト	1-gram	2-gram	3-gram	4-gram	5-gram
『論語』学而・為政	149.144	19.841	1.577	0.460	0.080
『論語』述而	174.953	16.930	1.319	0.381	0.280
『論語』顔淵	137.940	14.218	3.616	1.941	1.124
『礼記』緇衣	118.616	10.252	1.621	0.372	0.157
『礼記』坊記	126.215	17.505	3.600	1.844	1.044
『礼記』表記	144.362	13.892	3.520	1.222	0.281
『孟子』梁恵王	94.197	6.699	2.594	1.250	0.557
『孟子』公孫丑	104.374	6.857	2.099	0.873	0.437
『孟子』万章	104.850	9.520	2.846	1.406	0.563
『荀子』修身	95.056	7.940	3.254	1.769	1.036
『荀子』王制	118.829	7.262	1.869	0.618	0.220
『荀子』王覇	141.132	10.518	3.433	1.660	0.982
『荀子』正名	150.491	12.361	2.465	1.096	0.499
『荀子』非十二子	118.100	9.559	4.317	3.340	3.044
『荀子』礼論	126.642	12.112	4.254	0.825	0.315
『荀子』堯問	99.760	4.563	0.994	0.378	0.180
『墨子』尚同上	179.185	32.203	12.243	6.569	4.305
『墨子』尚同中	193.047	30.778	11.160	6.252	4.036
『墨子』尚同下	161.839	17.586	7.272	4.661	2.982
『墨子』兼愛上	199.343	35.361	13.652	8.384	5.688
『墨子』兼愛中	156.563	24.821	11.134	5.490	3.045
『墨子』兼愛下	155.187	19.664	8.478	3.525	1.825
『墨子』天志上	151.447	22.011	9.446	5.208	3.413
『墨子』天志中	170.860	21.567	9.312	5.866	4.181
『墨子』天志下	197.470	20.294	8.244	4.703	2.990
『荘子』逍遙遊	95.602	4.913	1.532	0.636	0.308
『荘子』斉物論	117.617	6.402	1.335	0.619	0.240
『荘子』天地	113.232	8.071	1.454	0.379	0.160
『荘子』天運	90.812	6.203	1.195	0.319	0.120
『荘子』庚桑楚	102.303	8.517	3.168	1.171	0.497
『荘子』天下	109.778	6.265	0.772	0.179	0.060
『呂氏春秋』有始覧	72.988	8.968	2.982	0.996	0.599
『呂氏春秋』孝行覧	99.741	7.351	2.655	0.940	0.361
『呂氏春秋』慎大覧	62.848	4.654	1.110	0.356	0.067
『韓非子』姦劫弑臣	158.553	12.077	4.420	2.323	1.243
『韓非子』顕学	134.810	10.080	3.413	1.360	0.561
『韓非子』孤憤	135.450	12.198	3.571	1.511	0.797
『韓非子』説難	148.446	10.120	3.140	1.603	0.884

表３：１～5-gramを対象とした*K*特性値の一覧

諸子書	論語	孟子	荀子	墨子	荘子	韓非子	呂氏春秋
異なり字種	1,221	1,242	1,313	894	1,468	1,179	1,483
共通字種	345	345	345	345	345	345	345
固有字種	168	140	189	47	307	120	242
延べ字数	12,599	12,599	12,599	12,600	12,600	1,2582	12,558
共有字数	9,868	9,837	9,359	10,523	9,346	9,404	9,012
固有字数	273	223	303	63	492	184	376

表4：異なり字数と延べ字数の比較

*K*特性値による分析

以下、求めた*K*特性値から、それぞれの学派別の特徴について述べる。

■　儒家

儒家は、後世の『史記』太史公自序に「夫儒者以六藝為法、六藝経伝以千万数」と文献（言葉）第一主義の学派としてその特徴が挙げてられているように、豊富な言葉が学派の信条である。

しかし儒家の言葉は当初から豊富だったのではない。孔子の言行録である『論語』の数値は、『論語』が短い文章を主体とする文章上の特徴もあって、諸子文献の中でのパターン化の度合いは高い。しかし、時代が降ると共に、より古典を積極的に取り入れ（或いは作り出し）、その解釈の形で儒家の言葉は増加する。

様々な言葉を取り入れて多様化するという傾向は『孟子』を一つのピークとする。文字種の表では『孟子』は『論語』『荀子』より数値が少ないが、これは『孟子』が『論語』や他の文献の言葉を積極的に取り入れ、『孟子』独自の主張のうち、後世に引き継がれなかったものが少なかったことを示しているとしてよい。その意味でも儒家は文献主義なのである。

『荀子』は*K*特性値の値にバラツキを示す。これは『荀子』の中に荀子の自著とされる篇と荀子後学の著作とされる篇が混在しているのが一因だが、『論語』から『孟子』にかけて減少する傾向にあった*K*特性値が、『荀子』の段階では数値が同程度か寧ろ増加の方向にあることからすれば、『荀子』学派が活動していた時期では、『孟子』の時よりも言葉が整理されつつあったとしてよい。

これは、『論語』の素朴な段階から、初期儒家（『礼記』緇衣・坊記・表記といったいわゆる『子思子』残篇）に至って新たな言葉が生まれ、『孟子』に至るまでの社会的激変を受けて更に言葉が増加し、それが前3世紀末の天下統一という時代の趨勢を踏まえて、新たなる言葉が生まれる一方、『孟子』以前では有効な言葉であってもこの時代では既に使用されなく（無意味な）なってきた言葉が存在したため、『荀子』では言葉に一定の淘汰が加えられたのである。

■　墨家

*K*特性値の結果を見ると、『墨子』諸篇は他の学派に比してどのgram数においても数値が高い。特に『墨子』と同じ前3世紀の成書である『荘子』や『荀子』と比較するとそれはより顕著である。これは墨家（ここではいわゆる十論を執筆した派）が使用していた言葉・文字が、他の学派に比してよりパターン化された言葉を使用していた事を示している。

また、文字種の表で他の諸子書と比較すると、『墨子』の使用字種数が少ないことが目立つ。『墨

子』が他の諸子書と共有する文字種は実に『墨子』全体の4割弱であり、他の諸子書と比較して1割程度多い。また、延べ字数の場合でも約84%を占め、これも他の諸子書より1割程度多い。

これは、墨家が「顕学」であったために、彼らの使用していた言葉が他学派で（肯定的否定的な意味合いで）使われたり、墨家が他学派の言葉を取り入れたりして言葉自体が共通化していった面もあるが、多学派の折衷的書物である『呂氏春秋』と比較しての数値の傾向の違いを踏まえれば、なにより墨家が使用していた言葉自体、他学派に比して貧困であったためであるといえる。

墨家は、戦国期に儒家と共に流行し「世の顕学」とまで称されたが、儒家がテキストを主体とする活動を行ったのに対し、墨家は当初より実践を旨とするという違いがあった。*K* 特性値や文字種の結果からも墨家が他学派のように新たな言葉・概念を積極的に創出する方面に力を入れず、寧ろ数少ない独自の言葉を重視していたことが見て取れるが、4 〜 5-gram での数値の高さからは、墨家がこれら貧困な語彙をパターン化し、それをスローガンとして活動していたことが伺える。

■ 道家

道家は *K* 特性値・文字種両表において、儒家と同様もしくはより数値が高い。従って多彩な言葉を駆使するのが道家の特徴であるといえる。荘子その人は孟子とほぼ同時代人とされるが、彼の自作あるいは彼の思想を最もよく反映しているとされる逍遙遊・斉物論では 1-gram が『孟子』とほぼ同様の数値を示す。但し、2-gramになると『孟子』よりも数値が低い。これは『孟子』よりも熟語レベルで新たな言葉を創出していることを示している。孟子がスローガンとしての自説を説く関係上、彼が重視する主張（＝言葉）を繰り返し使用するため、多少パターン化する傾向があるのに対し、荘子の言葉は、単一あるいは二字熟語を多用するため、大きな gram 数での *K* 特性値が低くなるのであろう。

『荀子』との比較では、やはり『荘子』の方が全般的に数値は低い。これは『荀子』が儒家の中である程度淘汰された言葉を使用するのに対し、『荘子』では自らの内面からあふれ出すイメージに対して個別に言葉を充ててゆくという違いがあると思われる。

この違いは文字種の面からでも見て取れる。「共有字数」は『荀子』とさほど変わらないものの、「固有字種」は儒家のほぼ倍、墨家の実に 6.5 倍という高い値であり、「固有字数」でも儒家の約2倍、韓非では約 2.7 倍、墨家に至っては約 7.8 倍も多い。この数値からも、道家が自らの固有の概念を新たな言葉に託して表現したかを明白に物語っているといえよう。

■ 雑家

『呂氏春秋』は、秦の宰相（相邦）呂不韋が諸学派を集めて編纂したという書物の性格上、そこには様々な言葉が内包される。

それは *K* 特性値では儒・道の中間程度の値になり、文字種の面でも「異なり字種」は最も多いが、「固有字種」の数は『荘子』よりは少ないという傾向を示すことからも明白であり、まさしく『呂氏春秋』は諸学を統合した雑家的な傾向を示しているといえよう。

■ 法家

『韓非子』は 1-gram での数値が墨家ほどではないが高く、韓非は使用する言葉をある程度絞り込んでいることが見て取れる。また 2,3-gram でも数値は『呂氏春秋』『荘子』に比較して高いが、『荀子』の一部と同様の傾向を示す。これは韓非が荀子の弟子であり、従って『荀子』後学でもあることを想起すれば、似たような傾向を示すのも道理であろう。

韓非は吃音であり、戦国遊子の「口から発する言葉」による主張展開が実質的に不可能であった。従って、彼にとっては文章が自身の主張を展開する唯一の武器なのであり、否応なく文章力を磨かざるを得ず、それが彼の文字種を絞り込む文章を構築する方向に向かったのである。

文字種の面から見ると、『韓非子』は「異なり字数」「固有字数」ともに儒家文献より少ない。

しかし墨家とは逆に「共有字数」は『荘子』『荀子』と同程度である。これは韓非が前3世紀によく使われた言葉で文章を組み立てつつも、その上に韓非オリジナルの言葉を洗練された形で展開するという文字使いをしていたことを示している。これは *K* 特性値の数値の高さ（よりパターン化された部分が多い）ことにも表れている。

おわりに

以上、N-gramモデルと *K* 特性値とを利用して、戦国期の諸子書の語彙量について分析を試みた。

本論での検討の結果、儒家・道家は多様な言葉を保持し、墨家は貧困、法家（韓非子）は洗練、雑家（『呂氏春秋』）はまさしく諸学を折衷した言葉を用いていることを数値的に明らかとした。

但し、儒家と道家では言葉の使い方に違いがある。儒家は文献主義という学派の特徴を踏まえ、前3世紀後半の『荀子』の時期には言葉を整理して絞り込む方向性を見せ始める。これは荀子後学でもある『韓非子』により顕著であるが、この推移は、戦国という乱世から天下統一へと移る時代を反映して、言葉も前3世紀後半以降新たな言葉（概念）の乱立が収まり、一定の整理の方向へと向かっていたことを反映しているとしてよい。

道家でも時系列で見た傾向は儒家と変わらないが、道家は学派の特徴である自らの内的イメージを表現するために、既存の言葉とは異なる新たな言葉を必要とし、そのために自学派オリジナルの言葉を多数創出したのであり、それが儒家との違いであるといえる。

しかし、道家特有の言葉を除くと、文章を構築する言葉の多くを他の学派と共有することは、道家の文献が記された前3世紀になると、学派を跨いだ共通の言葉（カールグレンの所謂「前3世紀の標準文語」）が形成されていたことも見て取れるのである。

本論での検討の結果、学派毎・時代毎の特徴をある程度明らかにすることができた。今後はより多くの文字資料（文献・出土）を対象として分析を行い、この傾向が果たして戦国期全般で成り立つのか、或いは学派（分派）・地域などで異なるのかについて検討を深めてゆきたい。特に今回は各諸子書の地域性については殆ど考慮に入れず分析を行ったが、現在、出土資料なども踏まえた各諸子書の成書地域に関する研究が蓄積されつつあり、将来的にはこれらを反映した言及も必要となろう [12]。

また、本論ではN-gramモデルを利用して分析を行ったが、上述のようにN-gramモデルではgram数が多くなればなるほどノイズデータが増加するため、共起頻度の数値をもとにデータを比較するという手法を採ることは難しくなる場合が多い。本論では5-gramまでを対象としたが、その場合でも多くの文献で *K* 特性値が1を、いくつかの文献では0.1を下回っている。そのような場合では、共起頻度が2のものすらほとんど見られず（本論で採り上げた『論語』学而・為政では5-gramの *K* 特性値が0.08であるが、延べ数999のうち、共起頻度1のものが991、同2のものが4であった）、語彙量の定量的変化を見るような場合では余り有効ではないと思われる。

従って、漢字文献をN-gramモデルを利用して定量的な手法で分析する場合、一般的にはgram数が少ない方が有益な結果が出る可能性が高い。しかし、高いgram数での共起が一致する場合、文献間で句法を共有するという可能性が高く、個々の文献を比較するという意味合いでは非常に有用なデータとなりえる。そのため、gram数については、分析の方法論に応じて適宜使い分けるべきだろう。

謝辞

なお、本論は科学研究費補助費・特別研究員奨励費：研究課題名「統計学的手法を利用した先秦文献の分類に関する研究─儒家文献を中心として─」（研究代表者：山田崇仁。課題番号6131）による研究成果の一部である。

注

[1] 「萬物」は『孟子』に一例見えるが、『孟子』と同時代にある程度の量が成書されていたと想定される『老子』、或いは道家に属する『荘子』に頻用されることからすれば、「萬物」が道家系の言葉であることは明白である。また、「萬物」は『荀子』以外にも『墨子』『管子』『呂氏春秋』でも頻用され、前3世紀には一般的な言葉として使用されていたことが伺える。

[2] 従来、漢字を「表意文字（Ideogram）」とする説が主流だったが、近年の言語学では、漢字が「意味だけではなく音も含んでいる」文字であることから、「表意文字」ではなく「表語文字（Logogram）」と呼ぶ。

[3] 『論語』は、孔子及び彼の弟子の言行録であるが、その成書年代に関しては様々な説がある。但し、『子思子』残篇である『中庸』『坊記』『緇衣』『表記』や『孟子』では『論語』が全面的に引用されており、少なくとも前4世紀には『論語』の祖本が遡ることは確実である。私見では前五世紀〜前四世紀交代期の成書を想定している。本論では、『論語』諸篇の中から初めの二篇と、成書が『論語』の中で相対的に降るとされる諸篇（季氏・堯曰など）を除いた諸篇から1000字程度のものということで述而と顔淵の二篇を選択した。

[4] これらが『子思子』の残篇であり、前4世紀前半のものであることは、武内義雄『易と中庸の研究』岩波書店　1943。後、『武内義雄全集』3　角川書店、1979所収。吉本道雅「緇衣小考」『東亜文史論叢』2号2004を参照。

[5] 『孟子』の成書年代に関しては、拙稿「『孟子』の成書年代について」『立命館東洋史学』27、2004を参照。

[6] 『荘子』のこの二篇については、荘周自作という伝統的見解を踏まえて、前4世紀後半に分類した。実際にはそれよりもやや降り、前4世紀〜前3世紀交代期の可能性もある。それ以外の諸篇については、外篇・雑篇から適当に選択した。

[7] 荀況の自著もしくはそれに近いもの、荀況後学のものを選択した。

[8] 韓非の自著とされる諸篇を選択した。

[9] 『呂氏春秋』序意篇に秦八年という紀年が記され、これは秦王政の八年（239B.C.）とされる。従って『呂氏春秋』の編纂はそれ以前の短い期間である。八覧から選択したのは、別な研究会で八覧を呼んでいる関係上、たまたま手元に電子テキストがあったためである。

[10] 墨家の代表的主張である十論から、時代的変遷を示す上中下全て揃っている篇を選択した。但し尚賢上については1,000字に満たないため、他のテキストに比して若干値が高くなっていると思われる。『墨子』十論の成書年代については、吉本道雅「墨子小考」『立命館文学』577、2002を参照のこと。

[11] 極悪氏は本名ではなくハンドルネーム（インターネットやパソコン通信で使用されるニックネーム）。
morogramは、SourceForge.jpで公開されている。具体的な使い方については筆者のWebサイト「睡人亭」を参照のこと。
morogram：http://sourceforge.jp/projects/morogram/
睡人亭：http://www.shuiren.org/
k-characteristics.plは、筆者が師氏とN-gramモデルの利用法について話している中で、師茂樹氏に作成していただいたものである。師氏には、篤く感謝する次第である。

[12] 筆者の知見の蓄積が少ないために確言はできないが、今のところは、地域性という条件で大幅にK特性値や文字種が変化するわけではないと考えている。それは、成書の時期（前3世紀）・地域（楚地）がほぼ重なると想定される『墨子』十論と『荘子』諸篇を比較した場合、本論で述べたように数値が大幅に異なっているというのが理由の一つである。

（研究ノート）

NGSM 結果のばねモデルによる視覚化

師　茂樹（もろ　しげき）

はじめに

本稿では、N-gram モデルを用いた複数文献間の確率・統計的比較分析（NGSM: N-Gram based System for Multiple document comparison and analysis）[1] において、従来あまり考慮されてこなかった出力結果の視覚化について、ばねモデルを用いた視覚化の試みを報告し、その有用性について検討する。

視覚化について

■　視覚化の必要性

視覚化（visualization）の元々の意味は、見えないものを見えるようにすること、もしくは見えないものが見えるようになることである。簡単な例で言えば、毎日の温度の変化を折れ線グラフなどで表現することは、一種の視覚化である。もっとも、コンピュータを使った視覚化は、折れ線グラフのように人間でも可能な単純なものよりも、大規模なものや複雑なもの（例えば WWW 全体のアクセス状況など）、あるいは抽象度の高いもの（文書間の概念的なネットワーク構造など）を対象にすることが期待されている。また、実際には細切れにハードディスク上に格納されているファイルを、最近の OS ではアイコンで表現し、操作できるようにしているが、これも一種の視覚化とされる（所謂デスクトップ・メタファ）[2]。

視覚化の目的にはいくつかあるが、本稿では特に、発想支援もしくは仮説生成支援の面に注目したい。仮説生成とは推論の一種（その一部は abduction とよばれる）で、先の折れ線グラフの例で言えば、例えばジグザグしながらも概ね右肩上がりの平均気温のグラフから、気候の変動を統一的に説明するための仮説（二酸化炭素排出量の増大が原因ではないか云々）を得る、というようなことである。土橋喜氏は、仮説生成による推論の特徴として、

> 仮説生成は蓋然的な推論方法であり、結論をそのまま真実であると認めることは適切ではない。しかし、ほぼ正しいであろうと認めることができ、ほとんど間違いのない命題を引き出す推論方法である。そのため結論が絶対的な確実性を持つものではなく、厳密には得た結果に対して検証が必要になる [3]。

と述べ、さらに、

> 学問の研究や日常の生活のさまざまな場面において、仮説生成が果たしている役割は

> 演繹推論に劣らないものがあり、その推論機能を無視することはできず、逆に支援することによって、人間の問題解決能力を拡大することが考えられる [4]。

と言う。

視覚化は、特にデータベースやネットワークが大規模化、複雑化している今日、通常では仮説生成が困難な対象から知識を抽出する方法として注目されており、近年は visual data mining と称されることもある [5]。

■ NGSM における視覚化

テキストデータベースの用途として検索が第一にあげられるが、検索はあらかじめテキストの内容やキーワード、表現上のパターンなどをある程度知っていないと有意な結果を得ることが難しい方法である。漢字文献のテキストデータベースが大規模化している現在、検索対象となるテキストの中で内容を把握しているものの割合は相対的に減り、大規模データベースとして十分に活用できない可能性が考えられる。

さらに言えば、重要だと思われる語の頻度を検索によって調べることはできても、頻度の高い語は何かを検索によって調べることはできない。言い換えれば、検索とは、研究者の先入観の範囲内でしか活用することができない方法なのである。

これを克服する方法として数理モデルや確率・統計的手法を用いたデータマイニングが注目されている。近藤泰弘氏は、コンピュータを用いる利点として、「現代人には通常認知できないデータの構造性や規則性探り出す」ことで、「現代人の古典語に対する「内省」(introspection)（語感）の欠如を補うことができ」る可能性を指摘している [6]。近藤泰弘・みゆき夫妻より端を発した N-gram モデルによる文献研究もまたそのひとつであり、現在、漢字文献の分野でも研究が蓄積され一定の成果をあげつつある [7]。しかし残念ながら、そのデータを人間が読む場合の方法や、分析結果の視覚化については、あまり考慮されてきたとは言えない。

これまで、N-gram モデルを用いた漢字文献の比較分析においては、

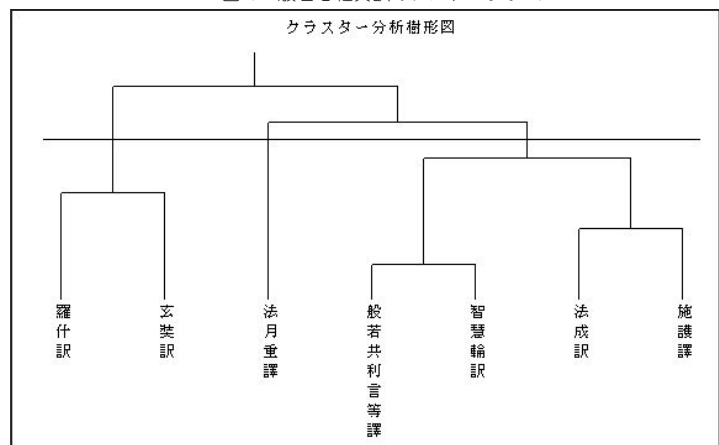
```
 4  雖願  2 （無，遊）
 4  雖見  2 （涅）
 2  雖盗  2 （梵）
24  雖然  2 （法，無，涅，海，戒，遊，
             二）
```

などのように [8] 文字や表などによって出力されたものを利用する場合が多かった。このような分析方法は、文献数が大きくなってくると人間が行う作業としては現実的ではなくなってくる [9]。

また、NGSM 結果をもとにクラスタ分析を行い、その結果を図 1 のようにデンドログラム（樹形図）として出力する方法も提案されている [10]。

この方法は有効であるが、近い個体どうしを順番にまとめていくため、例えば任意の 2 つの個体の距離を考える際には、直感的にわかりやすい視覚化とは言えない。また、注目する個体を、ユーザの要求に応じて変化させるようなインタラクティブ性もない。

先にも述べたように、漢字文献データベースが

図 1　般若心経異訳のデンドログラム

大規模化している今日、大量のNGSMデータから有意な仮説を引き出すためのひとつの方法として、さまざまな視覚化技術を検討するべきなのではなかろうか。

般若心経異訳のばねモデルによる視覚化の試み

■ NGSMのばねモデルによる視覚化

ここでは上記のような問題意識に基づき、『般若心経』の異訳間の関係を、NGSM結果に基づいてばねモデルによって視覚化することを試みる。

ばねモデルとは、「ノードがバネで結合されているという感じのモデル」であり、「バネはある長さのときエネルギーが最小になり、それより長くても短くてもエネルギーが大きくなる。エネルギーの総和が極小になるようにグラフを漸近的に変形していく」というものである [11]。これから示すように、先に述べたデンドログラムによる視覚化の弱点を克服する方法であろうと考えられる。

なお、『般若心経』異訳間のNGSMによる比較については以前論じたので [12]、視覚化の前処理についてはそちらを参照されたい。下の表は、3グラムで採取した結果をまとめたものの一部である（羅＝鳩摩羅什訳、玄＝玄奘訳、月＝法月重訳、般＝般若共利言等訳、智＝智慧輪訳、法＝法成訳、施＝施護訳）。

	羅	玄	月	般	智	法	施
一切世	0	0	1	0	1	1	0
一切大	0	0	0	0	0	0	1
一切如	0	0	0	1	1	1	1
一切法	0	0	0	0	0	1	1
一切苦	2	2	1	1	1	0	1
一切諸	0	0	0	0	0	2	0

この表に基づいて、各漢訳間の距離を求める。テキストA・B間の距離は、Aの変数 $A_1, A_2, A_3 \cdots A_n$ とBの変数 $B_1, B_2, B_3 \cdots B_n$ について [13]、以下の式で求める。

$$\sqrt{\sum_{i=1}^{n}\left(A_i - B_i\right)^2}$$

これによって求められる各漢訳間の距離は以下のとおりである（小数点第2位で四捨五入）。

	羅	玄	月	般	智	法	施
羅		18.5	31.6	31.1	34.2	36.6	38.6
玄			26.2	25.2	29.3	31.5	34.7
法				27.1	32.7	32.8	34.7
般					24.8	28.4	30.8
智						30.5	32.0
法							31.6
施							

これを各ノード間のばねの強さとする。

■ Javaアプレットによる描画

描画に際しては、Java 2 Platform, Standard Edition（J2SE）, v 1.4.2[14] 附属のデモ用アプレット [15]GraphLayoutが要求に近かったため、これを一部改造したものを用いた [16]。

アプレットに対しては、下記のようなタグ内に、「ノードa-ノードb/距離」の形式でパラメータを渡す。なお、以下の例では、描画の際の見栄えの関係上、距離を5倍にしてある。

```
<param name=edges value="羅什譯－玄奘譯/92,羅什譯－法月重譯/157,羅什譯－般若共利言等譯/155,羅什譯－智慧輪譯/171,羅什譯－法成譯/182,羅什譯－施護譯/192,玄奘譯－法月重譯/131,玄奘譯－般若共利言等譯/125,玄奘譯－智慧輪譯/146,玄奘譯－法成譯/157,玄奘譯－施護譯/173,法月重譯－般若共利言等譯/135,法月重譯－智慧輪譯/163,法月重譯－法成譯/163,法月重譯－施護譯/173,般若共利言等譯－智慧輪譯/123,般若共利言等譯－法成譯/141,般若共利言等譯－施護
```

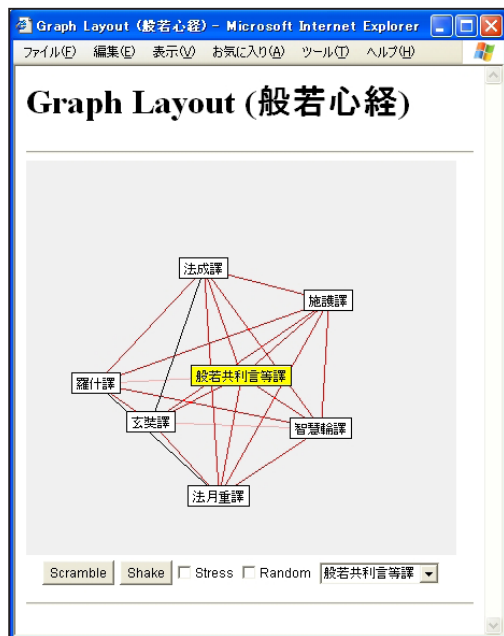



図２　ばねモデルによる視覚化（安定した状態）

図３　ノードを引っ張ったところ

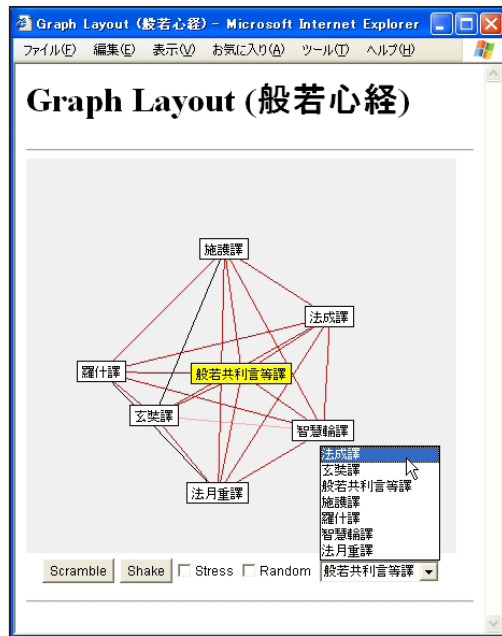



図４　固定するノードの変更

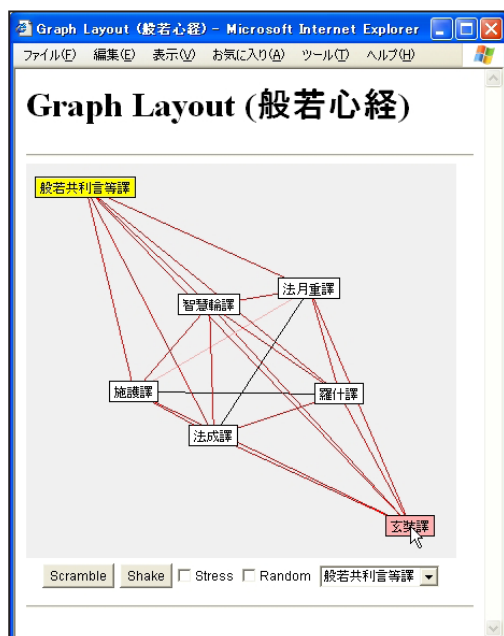
```
譯/153, 智慧輪譯-法成譯/152, 智慧輪
譯-施護譯/159, 法成譯-施護譯/158">
```

また、画面中央に固定したいノードがあれば、下記のようなパラメータを指定する。

```
<param name="center" value="般若
共利言等譯">
```

これによって図２のような画面が、ブラウザ中に得られる。各テキスト間の距離が、２次元で表示されている。

このアプレットでは、ごく簡単ではあるがいくつかのインタラクティブな操作が可能である。

一つは、各ノードをドラッグすることで移動させることができる点である（中央に固定したノードも移動できる）。ノードを移動させると、図２のような安定した状態になろうとする。ただし、固定されていないノードを移動させても固定されたノードは動かないので、ドラッグしている間は元の安定した状態にはならない（図3）。これによって、ノード間の関係をよりデフォルメされた

形で把握できる。

また、中央で固定されるノードは、任意のノードに変更することができる（図4）。これは、J2SE付属のGraphLayoutにはない機能であるが、先に述べた、任意のノードに注目してそのノードと他のノードとの関係を見たい場合には必要な機能であるため、追加したものである。

このようなインタラクティブ性はデンドログラムにないものであり、多角的な視点からの検討が可能になる。

■ 付論・ノードが大量の場合

筆者は、大規模な文献群もしくは大量の文献群に対するNGSM分析とその視覚化をひとつの目標としている。そこで、今回のアプレットに対して、ランダムに作った100個、500個、1000個のノードを表示させてみることで、スピードや視覚化としての有用性があるかどうかなどを検討してみた。

まず動作スピードであるが、筆者のPCのスペック（CPU：AMD Athlon XP 3200+、メモリ：1GB）では、100ノード（図5）ではまったく問題なく、500では少し遅さを感じ、1000ノード（図6）では現実的なスピードでは動かなかった。また、100ノード程度であれば、すぐにどのノードであるかが視認できるが、500以上になるとどれがどれだかわからず、視覚化の本来の目的を果たせないことがわかった。したがって、今回の方法では、ノード数の上限を（大雑把に言って）100程度までとするのが現実的であろう。

まとめと今後の課題

以上、雑駁ではあったが、NGSM結果のばねモデルによる視覚化について検討してみた。このような手法に関しては、動作スピードなどのベンチマーク的な部分を除いて、まだ明確な評価方法が確立していない。今回のようなばねモデルによる視覚化が、どの程度、仮説形成に貢献できるのかについては、現在のところ識者のご叱正を待つばかりである。

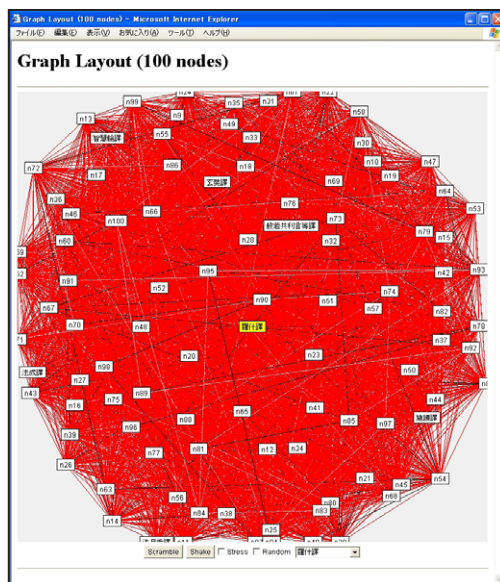
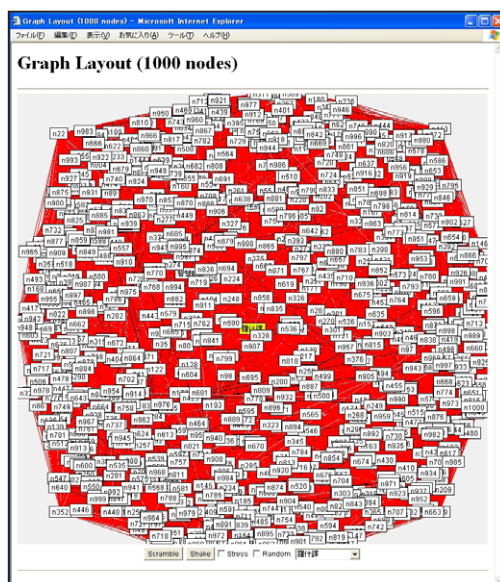



図5　100ノードの場合

しかし、文献間の複雑な関係を視覚化し、インタラクティブな作業を通じて知識を抽出するモデルやシステムが次々と開発され公開されている現在[17]、今後様々な分野でこのような取り組みがますます行われていくだろう。したがって、本稿の方法はごく簡単なものであったが、第一歩としては意義があるのではないかと考えている。

図6　1000ノードの場合

今後の課題としては、

- **拡大・縮小機能**
  現在は全体を見渡す視点しかないが、任意の部分を拡大するなどの機能は是非必要であろう。
- **フィルタリング機能**
  ユーザの要求や、注目しているノードからの距離などによって、表示するノードを動的に変化させることができれば、特に大量のノードがある場合に、視覚上の効果だけでなく、処理速度の向上も期待できる。
- **3次元化**

などを検討している。

謝辞

本報告は、科学研究費補助金・若手研究B「Nグラムモデルを用いたクラスタ分析による大規模漢字文献分析の基礎的研究」（課題番号15700215、研究代表者：師茂樹）および科学研究費補助金・基盤研究C「次世代中国古典文献データベース構築の基礎的研究」（課題番号14510494、研究代表者：村越貴代美慶応大学助教授）による成果の一部である。

注

[1] Ishii, Kosei. “NGSM and Cluster Analysis: Its Usage in the Digitization of Variant Texts in the SAT (Taisho Daizokyo Text Database).” *Proceedings of PNC Annual Conference and Joint Meetings 2002*. 2002.

[2] 増井俊之「情報視覚化の研究動向」（http://pitecan.com/articles/HIS/InfoVis/）

[3] 土橋喜『情報視覚化と問題発見支援　問題構造の可視化による仮説生成』（あるむ、2000）、p. 17。

[4] 土橋前掲書、p. 16-17。

[5] 増井前掲論文。

[6] 近藤泰弘「コンピュータによる文学語学研究にできること —古典語の「内省」を求めて—」（全国大学国語国文学会夏季大会シンポジウム「情報技術は文学研究をいかに変えるか」要旨、2001、http://klab.ri.aoyama.ac.jp/public/paper/20010602.pdf）

[7] 本誌第二号の特集等を参照されたい。

[8] 石井公成「仏教学におけるコンピュータ利用の現状」（『文学』隔月刊 第2巻第2号 2001年3、4月、岩波書店）など。

[9] 大量のNGSM結果から重要である可能性が高い要素を自動的に取り出す試みとして、師茂樹「Nグラムによる比較結果からの用例自動抽出 ——禅宗系の偽経を題材に」（『東洋学へのコンピュータ利用第14回研究セミナー』、2003）がある。なお、この方面に関しては、今後、主成分分析などの手法についても取り組む予定である。

[10] 師茂樹「Nグラムモデルとクラスター分析を用いた漢文古典テキストの比較研究——『般若心経』の異訳の比較を例に」（『京都大学大型計算機センター第69回研究セミナー「東洋学へのコンピュータ利用」』、2002）など。

[11] Tomihisa Kamada, Satoru Kawai. “An algorithm for drawing general undirected graphs.” *Information Processing Letters*, 31. 1989. に対する増井俊之氏のコメント（http://pitecan.com/bib/Kamada_DAGlayout.html）。

[12] 師前掲論文。

[13] 例えば羅什訳をA、玄奘訳をBとし、羅什訳の「一切世」の値をA_1とすれば、B_1は玄奘訳の「一切世」の値ということになる。

[14] http://java.sun.com/j2se/1.4.2/ja/download.html

[15] http://java.sun.com/applets/jdk/1.4/

[16] このプログラムについては、執筆現在、権利関係を確認していないため公開していない。確認がとれ次第、morogramのサイト（https://sourceforge.jp/projects/morogram/）より公開する予定である。

[17] 例えば、DocSpace（舘村純一「DocSpace: 文献空間のインタラクティブ視覚化」（『インタラクティブシステムとソフトウェア』4、近代科学社、1996）やNatto View（http://pitecan.com/presentations/DEWS98/NattoView/）、土橋前掲書の概念ネットワークなど。

Java 言語を用いて作成した N-gram 抽出プログラムについて

大谷　由香（おおたに　ゆか）

はじめに

２年前、龍谷大学淺田正博先生の大学院ゼミにおいてグループ研究を行っていた著者は、当時『漢字文献情報処理研究』第２号で紹介されていた、藤原滋氏作成のN-gram抽出プログラム「ngram」が、自らのグループ研究に活用できるのではないかと思い、このソフトウェアをゼミ内に紹介すると共に、オリジナルプログラムへの書き換えを検討した。

本稿では、当時Java言語を用いて作成したN-gram抽出プログラムである「norogram」[1]を紹介したい。試用の上、アドバイス等いただければ幸いである。

「norogram」について

「ngram」は、ソフトウェア自体がUnix環境にのみ適応するように作られたものであるからWindows・Macintoshなどのコンピューターで使用するには、まずOS上にUnix環境を構築することが必要となる[2]。著者の目的は、Unix環境を作らずともN-gram抽出が可能なプログラムを作成することであった。

■　Java プログラムを採用する利点

Javaでプログラムを作成した場合の利点の第1は、MacintoshでもWindowsでもUnix環境を構築せずにN-gram抽出ができるということである。すなわちMac OS XにはJavaを動かす環境が標準装備されているし、Windowsには雑誌付録のCDからインストール、もしくは以下のサイトからダウンロードすることで、簡単に環境を整えることが可能である。

- Sun Microsystems Java テクノロジ
  http://jp.sun.com/learnabout/java/

利点の第２は、JavaではUnicodeが標準となっていることである。Unicodeは他の文字コードに比べ対応文字が多く、また日・中・韓の漢字に汎用的に対応可能な文字コードであるから、漢字文献を扱うことを目的とする場合には、最も合理的な文字コードであるといえよう。

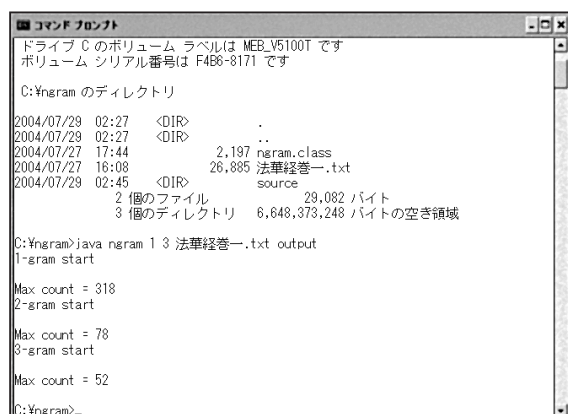
■　「norogram」の利点

現在公開されているN-gram抽出プログラムと比較したとき、「norogram」はそのプログラム上の機能として以下の利点が挙げられる。

第１は、対象となるテキストデータに含まれる、不要な文字・記号等をプログラム中で削除できる点である。「norogram」では、ひらがな・カタカナ・CJK統合文字・CJK互換文字以外は読み込まないよう設定されている。このため、半角や数字、アルファベットや句読点などの不要な文字、及び

中黒点や各種括弧類、？！；：などの記号がプログラム中で自動的に削除される。またスペースの削除と改行の無視についても「norogram」のプログラム中で行うことができる。これによって、N-gram 抽出する前段階としての煩雑なファイル処理が大方省けることとなる。

第２は、「ngram」は 1gram 抽出に対応できないが、「norogram」ではこれが可能 [3] な点である。これによって実際の研究へ応用しやすいプログラムとなったであろう。

「norogram」の使用方法

「norogram」は下記のサイトにて公開している [4]。

- **SourceForge.jp　プロジェクト：norogram**
  https://sourceforge.jp/projects/norogram/

サイトより「norogram」をダウンロードして解凍した後、以下の手順でプログラムを実行する。(□はスペースとする。)

1. ダウンロードした「norogram フォルダ」内に、対象となるテキストファイルを保存する。理論上、入力ファイルの文字数制限はないが、入出力の文字コードは UTF-8 形式のみである。
2. Java を動かす環境を立ち上げる。

```
コマンド プロンプト
 ドライブ C のボリューム ラベルは MEB_V5100T です
 ボリューム シリアル番号は F4B6-8171 です

 C:¥ngram のディレクトリ

2004/07/29  02:27    <DIR>          .
2004/07/29  02:27    <DIR>          ..
2004/07/27  17:44             2,197 ngram.class
2004/07/27  16:08            26,885 法華経巻一.txt
2004/07/29  02:45    <DIR>          source
               2 個のファイル              29,082 バイト
               3 個のディレクトリ   6,648,373,248 バイトの空き領域

C:¥ngram>java ngram 1 3 法華経巻一.txt output
1-gram start

Max count = 318
2-gram start

Max count = 78
3-gram start

Max count = 52

C:¥ngram>
```

Mac OS X 上では、
　Applications > Utilities > Terminal
Windows 上では、
　プログラム＞アクセサリ＞コマンドプロンプト
を立ち上げる。

3. プログラムを「norogram フォルダ」内で始動するため、cd □と入力した後に、「norogram フォルダ」をドラッグし、改行する。
4. 下記のコマンドを入力し実行する。
   java □ norogram □〈開始 gram 数〉□〈終了 gram 数〉□〈入力ファイル名〉□〈出力ファイル名〉
5. リターンキーを押す。
6. 「norogram」が動き始める。

（左下の画像は Windows 上で「norogram」を始動したもの。『法華経』巻 1 を題材として、1gram から 3gram までを指定して作動させた。）

N-gram 抽出結果は、1gram から順に完了次第、「norogram フォルダ」内にテキストファイルで出力される。

例えば『法華経』巻 1 を 3gram 解析した場合、出力テキストファイルの一部を挙げれば以下のよ

1	是阿羅	3
8	阿羅漢	3
1	羅漢諸	3
1	漢諸漏	3
1	諸漏已	3
2	漏已盡	3
1	已盡無	3
1	盡無復	3
1	無復煩	3

うになる。

右端の数字が頻度、左端の数字が gram 数を表す。つまり『法華経』巻 1 中に、「阿羅漢」という 3 文字の組み合わせは 8 回、「漏

己盡」という３文字の組み合わせは２回、その他の３文字の組み合わせは１回みられるということがわかる。

この解析結果はその頻度が全て１になるか、終了gram数までグラム数を増やしてフォルダ内に蓄積されていく。

「norogram」の今後の課題

本稿では、「norogram」の紹介とその使用方法について述べるに留まったが、もちろんこのプログラムは研究のためのツールの１つであり、これから得られるN-gram抽出結果を基にして、石井公成氏の提唱するNGSM分析を行うといった応用のないことには研究としては成り立たない。

このように研究に応用するとき、その都合上、複数のテキストデータを一気に解析していく手段が必要とされることが予想される。「norogram」が複数のテキストデータを一気に解析していくことが可能となれば、かなり実用的となるであろう。本プログラムでは、Javaのvectorクラスを使用した。複数文献を同時に処理するためには、vectorクラスの要素を追加することによって可能になるのではないかと思われる。

また「norogram」はgram数が増すごとに動作速度が遅くなる。これについても何らかのプログラム改良によって解消することが可能であろう。

以上のように「norogram」は、研究に応用する上では課題の残るプログラムであり、またプログラミングに精通していない筆者の作成したものであるから、不具合や改良の余地が大いにあると思われる。「norogram」のプログラムは、改良・転載等が自由にしていただけるよう、先のサイトにて公開している。皆様に試用していただき、知恵を拝借することによって、よりよい研究ツールとなることを期待している。

注

[1] この名称は、作成当時に既に発表されていた師茂樹氏作成の「morogram」に敬意を評し、これをもじって仲間内で使用されていたものである。今回発表するに当たって師氏の許可を得てそのまま使用させていただいた。

[2] Macintoshは、Mac OS X以降の場合はUnixベースであるから、最初からUnix環境が整っている。しかし筆者の力量不足のためMac OS X上でソースコードをコンパイルすることができなかった。

[3] 師茂樹氏はN-gramをPerlスクリプトとC言語によって書き直した「morogram」を作成することにより既に頻度１の検出を可能としている。

[4] ちなみに「morogram」も現在は同じくsourceforge.jpにて「プロジェクトmorogram 」として公開され、共同開発されている。
https://sourceforge.jp/projects/morogram/

実践レポート

今年は「電子辞書」をテーマに採り上げた。
昨今、電子辞書が学生の間でごく普通に利用されている。英語や中国語の本格辞書を丸ごと搭載していることに加え、国語辞典等との連動検索や音声出力など、従来の辞書にない使い方を可能にしている点が人気の秘密のようだ。一方、このような電子ツールを学習に利用することについて懐疑的な教員もまだまだ多い。
電子辞書は中国語学習にどのような影響を与えるのか？　授業者にとって理想の電子辞書とは？
本特集は、これらの疑問に応える、研究会なりのアプローチである。

日本中国語 CAI 研究会について

本会（会長：田邉鉄北海道大学助教授）は「コンピュータ援用の授業方法を中心とした中国語教授法の研究・開発・普及を推進し、同時に教員・研究者・ソフトウェア開発者の交流をはかる」ことを目的とし、1996 年 11 月に発足した。
会員による研究発表・実践報告の場として、例会（年 1 ～ 2 回）、総会（秋、年 1 回）を開催するほか、常時メーリングリストで情報意見交換を行っている。
参加を希望される方は、中国語 CAI 研究会 Web　**http://moli.cims.hokudai.ac.jp/~ccai/**　を参照していただきたい。

Contents

電子辞書は授業で使えるか？

オンライン討論から

日本中国語CAI研究会　田邉　鉄（たなべ　てつ）

✎ はじめに

中国語CAI研究会では、最近相次いで発売された中国語電子辞書が、学習ツールとして有効かどうかを検証するため、特設掲示板[1]でオンライン討論を行った。メーリングリストで田邉鉄会長が呼びかけたところ、5月から8月の間に8名が参加し、50余りの発言があった。

本稿はこの討論の中から、特に興味深いと思われる話題を取り上げて、田邉が構成したものである。発言の引用にあたっては、発言者の意図をなるたけ配慮するよう心がけたが、発言の全体や議論の流れそのものを忠実に反映したものではないことをお断りしておく。実際の議論については、現在も運用中の掲示板を参照頂きたい。

なお、引用はゴシック体で、発言者は引用の後に（　）でそれぞれ表記した。

✎ 学生が例文を読まなくなる？

「電子辞書は学生が例文を読まなくなるから反対」という意見がありますが、みなさまのお考えはいかがでしょうか？（あやうく「電子辞書買うな」みたいな指導を強要されかけた）私自身は紙の辞書でも学生はたぶん最初の方にある意味を見るだろうし、大部分の学生は例文は読まないから同じだ、という考えです。（金子眞也）

この他、中国語ではないが、手軽に引ける電子辞書を使うと授業中に辞書引きするクセがつき、しっかり辞書を引いて予習しなくなる、という新聞記事も紹介された。

こうした問題点の指摘に対しては、「例文を読まない学生は、紙だろうと電子辞書だろうと読まない」という意見が続出した。

電子辞書は学校が用意するCALL設備とは異なり、個人で購入・使用するものである。それをどこでどのように使って、授業に役立てる（あるいは手抜きをする）かは、あくまでも学生個人の自覚の問題と言える。

ただ、紙、電子を問わず、「辞書との付き合い方」のようなものが変化しつつある中で、学習過程に辞書をきちんと位置付けて考える必要があるという問題提起もなされている。

✎ 辞書引き指導のあり方

根本的な問題として、辞書引きは学習過程においてどんな作用があるのでしょう？外国語学習において辞書を引くことは自明だと思っていたが故に、真剣に考えたことがありませんでした。（清原文代）

辞書の記述を有効に活用するためには、具体的な課題を与えねばならないと考えています。「先生、ここの意味がわかりません！」「辞書を引け！」では、学生は、訳語しか見ませんよね。

具体的には、次のような課題を考えています。

「“結婚” は、動詞＋目的語の構造だから、結婚相手を示す名詞句は、“結婚” の直後には置けず、前置詞をつけて、動詞の前に置かないといけない。どの前置詞を使わないといけないか、辞書で確認しなさい。」（山崎直樹）

ふと思いついたのですが、類義語の比較などもおもしろいかもしれません。最近の辞書は類義語解説に力を入れていますから、その解説を見つけて読む機会になります。

電子辞書ならではの練習というのもありうるでしょうね。例えば、後方一致検索を使って「＊心」の単語を集めて比較するとか。検索が容易なことを利用して、日中同形語を探させて、その中から日本語と意味や用法の違うものを指摘させるとか。（清原文代）

筆者も「辞書を引くことが自明」と思い込んでいる（た）クチである。授業中に「辞書の引き方」を指導することはあっても、「辞書の使い方」を指導するという発想は全くなかった。辞書には訳語以外にもたくさんの情報が含まれているが、それが学習過程でどのような意味をもっているか、どう使えば役に立つのか、ということを教えていないのだから、「例文を読まない」のも無理からぬことであろう。

ここでの議論は、紙か電子辞書かではなく、中国語授業における辞書一般の位置付けについてであるが、電子辞書の活用というテーマにも重要な示唆を与えている。授業で具体的な課題を提示するということは、辞書を使う必要性、必然性を示すということでもある。清原氏が示した後方一致検索を利用した、電子辞書ならではの練習という例は、従来の辞書では利用することの難しかった「情報」を利用することにより、語への理解を深められるであろうことが容易に想像できる[2]。

ここで挙がった例のような課題を蓄積し、授業の進行に合わせて適切に提示する、それが本当の意味で「授業で辞書を使う」ことになるのではないかと思う。

✎ 理想の電子辞書

中国語電子辞書について、多くの利用者、特に教員が不満に感じるのは「辞書が選べない」ことである。昨今の電子辞書ブームは、市販辞書の中身をフルに収録した機種の登場によるところが大きいが、中国語の場合、ほぼ例外なく小学館の中日・日中辞典の旧版しか収録されていない。

仮に理想的な中国語電子辞書があるとすれば、どのような姿をしているだろうか。

検索機能としては、

- 例文を含めた全文検索
- 複数辞書の串刺し検索
- 漢字単位の前方一致・後方一致検索
- 漢字の音訓検索
- 四角号碼による検索

収録辞書は、

- 小学館中日第 2 版
- 小学館日中第 2 版
- 講談社中日第 2 版
- 白水社中国語辞典
- 東方中国語辞典
- 大漢和辞典
- 現代漢語詞典
- 漢語大詞典
- 漢語大字典

まぁ、ありえない夢のような話なのですが、たまには夢を語るのもよいかと…（清原文代）

日本語の類義語辞典と連繋した電子辞書がほしいですね。ぼちぼち類義語辞典・シソーラスの類も出版され始めていますから、それらを収録して、ジャンプ機能を工夫することで、だいぶん良くなるのではないかと思いますね。

さらにシソーラス機能を作り込んで、各辞書の

収録語彙の意味による分類検索・一覧機能、なんてのも面白いかも。（千田大介）

電子辞書の特徴を活かした機能として、複数辞書の串刺し検索がある。同じ語の解説や例文を複数の辞書について一覧できれば、学習、研究に大いに役立つであろう。ただ、電子化することによって紙の辞書が売れなくなるのではないか、という出版社の不安や、コンテンツ代が価格に直結する製品の特徴もあり、実現には困難が予想される。例えば目先のロイヤリティだけではなく、電子化することで紙の辞書の価値も上がる、というような Win-Win 型のビジネスモデルが構築できなければ、夢の実現は不可能であろう。

✎ 学生用電子辞書の適正価格

学生向けの電子辞書の適正価格ってどれくらいなんでしょう？個人的にはやはりせめて１万円代にならないと売れないような気がしますが。

価格を下げるためには、ハードウェアのスペックを落とし、収録辞書を減らすことが必要なのでしょうが、そのバランスが難しそうです。ハードウェアについてはよくわかりませんけれど、収録辞書の最低限の構成なら、

- 中日辞典１冊
- 英和辞典１冊
- 和英辞典１冊

というのもありだと思います。（清原文代）

現在のロイヤリティ（つまり出版社や著者のマージン）が変わらないなら、これ以上安くなると本当に改訂の費用が捻出できません。（現状でもかなり困難です。）もう一歩、二歩安くなればもっと売れるのは事実でしょうが、出版社としては乗り気になれない状況です。（岩野忠昭）

紙の辞書の場合、形も大きく語彙数も膨大な辞書（たとえば大漢和）とコンパクトなポケット辞書（デイリーコンサイス）のような棲み分けが存在しますよね。もちろん、その間に中辞典や学習初級辞典などもあるわけで…。

電子辞書の場合、こういった収録語彙数・例文数の多寡で差を付けた棲み分けというのはあり得ないんじゃないかと思います。収録辞書の数によって、廉価版を作るという発想は可能だと思いますが、例えば収録辞書が半分になった時、どの程度定価が下がるのでしょうね？　あまり期待できないような…。（岩野忠昭）

中日、日中辞典を２冊買うと結構な値段である。また、筆者の勤務先では英作文の授業で英英辞典や類語辞典も活用されることなどから、英語辞書が充実していさえすれば、２万円台後半の価格は必ずしも「高い」とは思われないようではある。もっとも、完全に紙の辞書の代わりに使っているという学生は少なく、電子辞書を使う学生は紙の辞書も買って、それなりに使い込んでいる、ハイエンド系の学生が多いのだが。

もちろん、「たくさんの辞書が入っていてお得」とはいえ、２～３万円という金額は、学生にとっては大金である。そこに収録辞書を減らし、廉価版を作る、という発想が生まれる余地がある。

もっとも、収録辞書を減らすといっても、メイン・コンテンツである中国語や英語系の辞書をそのままに、他の「おまけ」的な辞書（これらはおそらく相当安く買い叩かれているだろう）を減らすことで、新たな需要を喚起するほど値段を下げられるかどうかは、疑問である。買わない学生は５～６千円安くなった程度では買わないだろう。身の回りで電子辞書を購入した学生は、おおむね紙の辞書も熱心に引くであろう「優等生」が中心である。それ以外の層の学生にとってみれば「便利」だけでは電子辞書を購入するモチベーションとしては弱いのではないか。

もちろん、学生の経済的負担は低いに越したことはないので、例えば初級学習辞書を英語電子辞書に加えた、本当の「第二外国語モデル」や、複数の中国語辞書を搭載した「専門課程モデル」などを熱望する。いずれにせよ、学生の中国語学習

に必須のツールとなるべく、収録辞書の選択や学習機能の追加などの工夫が必要であろう。

✎ 電子辞書ならではの「見せ方」

電子辞書は、紙の辞書と同じ「見せ方」をする必要はないんじゃないかと思っています。動画で、辞書製作者の意図を反映したプレゼンを自動的に行う、とか。僭越ながら素案を…。

1） 利用者が“拍”を引く
2） 用例の一覧（下記参照）が出て、「どれにいちばん近いですか？」というダイアローグが出る

(a) 拍卓子大罵、拍巴掌
(b) 拍電影、拍照片
(c) 拍電報給他

3） 例えば、(a) を選ぶと、いきなり次のようなアニメが始まる。

── 机をバンバンたたいて怒鳴っている情景
── 拍手している情景
── ごみを叩き落としている情景
── 鳥がはばたきしている情景

4） やっと、日本語による“拍”の語釈が提示される

こんな疲れる辞書、誰も使ってくれませんね。
（山崎直樹）

山崎氏の指摘している通り、電子化によって「見せ方」の工夫はいくらでも広がる。単語の説明をする時、「これはアニメーションで見せた方がわかりやすいだろうなぁ」と思うことはよくある。「それを何が何でも言葉で説明するのが辞書だ」という言い分もわかるが、絵や音、動画などで補助すると、簡単に分かるような語があるのも事実である。今のところ、音声が聞ける機種が増えていることを除けばこの方面で目立った工夫はないが、中国語以外の電子辞書では、カラー液晶を搭載したものも出始めている。メモリの大容量化や、処理速度の一層の高速化、液晶ディスプレイの高解像度・多色化などの技術的課題が解決すれば、マルチメディア化が進むことも考えられる。

もちろん、マルチメディア・コンテンツは電子辞書用に新たに開発する必要があり、それは100% 本体価格に反映されるだろうから、凝ったものは難しいだろうが、教室でちょっと説明が面倒だな、と思うような語に限定しても、かなり使えるのではないかと思うが、どうだろう。

また、学習者向けに簡易ドリルを付け加えるのは、それほど難しくないだろう。見せたいものだけを選んで見せることのできる電子辞書の特徴を活かした、検定対策単語ドリルなどは有効であるように思う。実際、英語では電子辞書と TOEFL、TOEIC 試験対策ドリルを組み合わせた機種の売れ行きが好調だと言う。

✎ おわりに

討論では、ユーザインタフェースや検索機能についても議論されたが、本誌清原氏のレビューに詳しいので、重複を避けるため割愛した。

今回は主として、授業での利用を想定した工夫についてまとめるつもりだったが、市販の電子辞書に対する不満があちこちに顔を出すことになってしまった。掲示板に掲載された不満や希望については、ぜひともベンダからの意見も伺いたいと思う。この場で、各社担当者の掲示板への参加をお誘いしておく。

注

[1] http://moli.cims.hokudai.ac.jp/edic/
ゲスト利用も可能なので、会員以外でも興味のある方はぜひ参加して頂きたい。

[2] もちろん、「引き方」にも習熟するであろうから、「紙の辞書と電子辞書、両方引き方を教えるのはムダが多い」という至極もっともな意見に対する回答でもある。

中国語電子辞書機能比較

清原　文代（きよはら　ふみよ）

✎ 1. はじめに

1997年にSONYからDD-CH10（現在は販売終了）が発売されて以来、しばらく音沙汰がなかった中国語電子辞書が去年今年と相継いで4つのメーカーから発売された。

CASIOのXD-7300とセイコーインスツルメンツ（以下SII）のSR-T5030については、前号で田邉鉄氏がレビューを書いておられる[1]。本稿ではその後発売されたCanonとSHARPのものを加えてレビューをお届けする。

✎ 2. 機種一覧

2004年7月現在で各メーカーの発売中、または発売予定の機種は以下の通りである。

メーカー名	機種名
CASIO[2]	XD-H7300・XD-H7310 XD-L7350・XD-L7360 XD-M730
Canon[3]	Wordtank V70
SHARP[4]	PW-A8500
SII[5]	SR-T5030 ST-T7030

（配列はメーカー名の五十音順）

1つのメーカーに複数の機種がある場合でも、機種の違いは基本的には音声の有無や中国語以外の収録辞書の相違を表している。本稿では筆者が所有する各メーカー1機種ずつを中心に項目別に比較していく。

本稿で主にレビューする機種

- XD-H7310（CASIO）
- Wordtank V70（Canon）
- PW-A8500（SHARP）
- SR-T5030（SII）

✎ 3. 中国語辞典

3.1 収録されている辞書

搭載されている中国語辞典についてはほとんどバリエーションがないと言ってよい。4機種とも小学館の『中日辞典』『日中辞典』の初版を搭載している。唯一PW-A8500だけが小学館の辞書に加えて、朝日出版社の『はじめての中国語学習辞典』を収録している。

小学館の辞書は初版が出版されてからすでに10年以上が経過しており、現在は第2版になっている。電子辞書の方も第2版の搭載が望まれる。また、英語の電子辞書の場合は複数辞書の搭載はごく普通のことであり、中国語の電子辞書においても今後辞書の種類が増えることを期待する。

3.2 中日辞典の検索語入力

4機種とも拼音・日本語読み（単漢字）・部首や画数による検索語の入力をサポートしているが、それぞれ操作方法に相違があり、同じ小学館の辞書であっても使い勝手がかなり異なる。

● XD-H7310

『中日辞典』の初期画面は検索語入力方法の一

覧で、そこから使いたいものを選び、決定ボタンを押して入力画面に入る。検索に入る前に必ず入力方法の一覧画面を経なければならないのは些か面倒である。

● Wordtank V70

初期画面は拼音入力、他の検索法は検索法ボタンを押して表示させる。

この機種の最大の特徴は付属スタイラスペンによる簡体字の手書き入力ができることである。認識できるのは 8651 字。

高校で漢和辞典を引く機会が減っているのか、中級上級になって教材から拼音がなくなると、部首引きができないため途方にくれる学生がいるが、手書き入力であれば、それが良いことかどうかは別として、漢字の構造に対する知識がなくとも、とにかく簡体字が書ければ検索することができる。

● PW-A8500

小学館『中日辞典』の初期画面では拼音と日本語読みの入力欄が表示される。部首や画数入力画面に行くにはカーソルを移動させて決定ボタンを押す必要がある。なおこの機種だけが「熟語検索」を使って単漢字ではなく単語単位の日本語読み入力と部首引き [6] が可能である。

朝日出版社『はじめての中国語学習辞典』は拼音検索のみ。

全辞書検索を使うと、小学館『中日辞典』と朝日出版社『はじめての中国語学習辞典』の拼音による串刺し検索が可能。

● SR-T5030

初期画面で拼音・日本語読み・部首読み・部首画数・総画数の入力欄が示される。一番単純で且つ使いやすく感じる。

3.3 声調の入力方法

4 機種とも拼音入力の際に声調符号はなくてもよいが、声調を入力することも Wordtank V70 以外は可能で、その方法は三者三様である。

● XD-H7310

アルファベット＋声調を表す数字。

（例）hǎo ＝ hao シフト 3

これが一番速い入力法である。

● Wordtank V70

声調の入力はできない。

● PW-A8500

声調符号をつける母音を複数回打つ。

（例）　a を 1 回タイプすると→ a
a を 2 回タイプすると→ ā
a を 3 回タイプすると→ á
a を 4 回タイプすると→ ǎ
a を 5 回タイプすると→ à

声調を入力しようとすると打鍵回数が多くなるのが面倒である。例えば hǎo を入力しようとすると、h aaaa o と打鍵しなければならない。

● SR-T5030

キーボードに四声キーがあって、そのキーを押す度に 1 声→ 2 声→ 3 声→ 4 声と表示が変わっていく。専用キーがあるため声調入力の方法が直感的にわかりやすい。

（例）hǎo ＝ h a　四声キーを 3 回打鍵　o

ただ問題は上記の例からもわかるように PW-A8500 と同様に打鍵回数が多くなることと、声調符号を付ける母音のすぐ後に四声キーを押さなければならないことだ。例えば "好" を検索しようとして、ついうっかり hao と入力した後に四声キーを 3 回打鍵してしまうと、画面には haǒ というありえない拼音が表示されてしまう。声調符号をつける母音を選ぶ規則はさして難しいものではない。機械に判断させてしまえないものだろうか。

3.4 検索結果の表示

一覧性については、電子辞書は紙の辞書にどうしてもかなわない。とはいえ小さな液晶画面の中にいかに多くの情報を見やすく表示するか各社工夫している。

3.4.1 検索結果リスト

XD-H7310とSR-T5030が画面分割式を採用し、上段の画面にリスト、下段の画面にカーソル位置の本文の冒頭部分が示される。基本的な意味をぱっと見たいときには便利だ。

3.4.2 例文の表示

例文の表示について以下の二つの方法があり、それぞれ一長一短である。

A: 意味と例文が同時に表示される。
B: 例文は表示されず、意味だけが表示される。例文を見るにはボタンを押して表示させる。

PW-A8500とSR-T5030はA式、XD-H7310とWordtank V70はB式[7]である。

A式では、意味項目の多い単語だと何度も画面をスクロールさせていかなければならない。スクロールしているうちに最初に何が書いてあったかわからなくなる学生もいるであろう。

それに対してB式は単語がどんな意味を持っているかを概観することができ、一覧性においてA式に優るが、それと引き替えに例文を見るために新たに別のボタンを押すという手順が1つ増える。これが学生の学習にどう影響するのか、私の担当するクラスでは中国語電子辞書を持っている学生がまだ少なく、今のところ何とも言えない。電子辞書に対する批判の1つに学生が例文を読まなくなるという意見があるが、私の経験では紙の辞書でも電子辞書でも例文を読まない学生は読まない。紙の辞書であれ電子辞書であれ例文を読むことの必要性を根気よく指導するしかないと考えている。

3.5 逆引き・ワイルドカードを使った検索

いずれの機種も操作方法は異なるものの、『中日辞典』の逆引きやワイルドカードを使った検索が可能である。これらの機能は拼音をうろおぼえの時やディクテーションで単語の一部しか聞き取れなかった時などに大いに力を発揮する。

XD-H7310では逆引きのための専用入力欄を設け、拼音入力欄でもワイルドカードを使った検索が可能だ。他の3機種は逆引き専用の入力欄は設けていないが、拼音入力欄でワイルドカードを使った検索ができ、*（任意の1文字以上にマッチ）を使って逆引きをすることも可能だ[8]。

しかし大きな問題がある。これらの検索が漢字単位ではなく拼音のアルファベット単位なのだ。例えばhaoで終わる単語を検索するために逆引きするとzhao chao shaoで終わる単語もリストアップされてきてしまう。

ワイルドカードによる検索も同様で、？（任意の1文字にマッチ）は拼音のアルファベットの1文字を指し、漢字1文字を指定することができず、中国語検索においては実質的にあまり役に立たない。*（任意の1文字以上にマッチ）[9]も、拼音のアルファベット単位である。

多彩な検索方法を提供できることは電子辞書の強みで、ワイルドカードを使った検索などはその最たるものである。しかし中国語の検索においては現時点ではその特徴を生かし切っていない。拼音のアルファベット単位だけではなく、漢字単位のワイルドカード検索ができるように改善を強く希望する。

3.6 囲み記事

小学館の辞書の特長の1つは囲み記事が充実していることである。4機種とも囲み記事を収録しているが、表示のさせ方はかなり異なる。

● XD-H7310

囲み記事の項目一覧は無し。

辞書本文では囲み記事は「囲み」「文法」と表示され、ジャンプ機能を使って表示させる。ジャ

ンプ機能を使うため当該の箇所までカーソルを移動させた上で決定ボタンを押す必要がある。

用例を表示しているときには、画面に「囲み」の表示が見えていても、一旦用例表示モードを解除してからでないと囲み記事を表示させることができない点が不便。

● Wordtank V70

シフト＋中日ボタンで『中日辞典』『日中辞典』の囲み記事と世界の人名・地名[10]の項目一覧を表示させることができる。

辞書本文中では、囲み記事は「囲み」「文法」と表示され、ジャンプ機能を使って表示する。当該の箇所までカーソルを移動させ、決定ボタンを押す必要がある。

● PW-A8500

『中日辞典』『日中辞典』のそれぞれの初期画面から囲み記事の項目一覧に行くことができる。

辞書本文中では､囲み記事は「解説」と表示され､解説ボタンを押し、更に決定ボタンを押して表示させる。画面上に「解説」が１つしかなくとも、決定ボタンを押す必要があるのは少し面倒だ。

● SR-T5030

囲み記事の項目一覧は無し。

辞書本文中では囲み記事は「囲み」と表示され､例文・解説ボタンを押すと、画面が２分割され、囲み記事の冒頭数行が下段画面に表示される。決定ボタンを押すと囲み記事の全文が表示される。４機種の中では一番使いやすいと感じる。

✎ 4. 中国語会話集

4.1 収録されている会話集

（右上表参照）

XD-H7310	ひとり歩きの中国語自由自在（JTB）
Wordtank V70	ひとり歩きの中国語自由自在[11]（JTB）
PW-A8500	らくらく旅の中国語（三修社）
SR-T5030	中国語会話とっさのひとこと辞典（DHC）[12]

4.2 会話集の検索機能

XD-H7310とWordtank V70は場面別の項目一覧をたどっていく形式。キーワードによる検索はできない。

PW-A8500は場面別項目一覧の他に、日本語キーワードによる検索ができる。

SR-T5030は場面別項目一覧に加えて、検索機能が充実しており、日本語検索･中国語検索[13]･日本語キーワードによる検索・中国語キーワードによる検索ができる。元になった書籍にも日本語索引・中国語索引があるが、SR-T5030では日本語検索・中国語検索においてそれぞれ複数の単語によるand検索が可能で、電子辞書の特長を生かしていると言える。

4.3 会話集の発音表記

XD-H7310は拼音及び声調符号を付けたカタカナ、Wordtank V70は声調符号を付けたカタカナのみ。PW-A8500はカタカナのみ（声調符号無し）、SR-T5030は拼音及びカタカナ（声調符号無し）。

４機種ともカタカナによる発音表記を採用しており、抵抗を感じる教員も多いであろうが、元になっている書籍が教室での利用を想定したものではなく、簡便に会話文の発音を紹介するという点でやむを得ないのであろう。しかし、教育的観点から見れば、必ず拼音を併記して欲しい。

XD-H7310とWordtank V70は同じ『ひとり歩きの中国語自由自在』を収録するが、上記のようにXD-H7310は拼音を併記し、Wordtank V70はカタカナのみである。XD-H7310の取扱説明書によれば、原著には拼音はなくCASIOが独自に追加したものであるという[14]。中国語に限らず電子辞書に収録されている辞書は紙の辞書をそのまま電子化したものが大半という状況下で、会話集とはいえわざわざ原著に存在しない拼音を

追加したことは、学習者への配慮及び電子辞書独自のコンテンツという点で注目すべきである。隴を得て蜀を望む、設定によってカタカナ表記を消すことができれば更に良いと思う。

✎ 5. 音声

紙の辞書にはできない電子辞書ならではの機能に音声がある。拼音字母と実際の発音の関係がしっかり定着していない初級レベルの学習者にとって非常にありがたい機能である。

中国語音声を収録しているのは、XD-L7350・XD-L7360・Wordtank V70 の３機種である。

XD-L7350・XD-L7360 は、中日辞典については親字約 13000 字・熟語約 1300 語、会話集は例文約 2300 の音声を収める。中日辞典も会話集もネイティブスピーカーの発音を録音したものである。

Wordtank V70 は、会話集はネイティブスピーカーの発音を録音したものだが、中日辞典の親字と熟語の発音はソフトウェアによる合成である。そのため XD-L7350・XD-L7360 とは異なり、中日辞典の見出し語全ての発音を聞くことができるのだが、一方で問題点もある。音声を合成するため２音節以上の単語は音節と音節の間に一瞬の間が空くことがある。また、“不”と“一”の声調変化には対応しているが、３声＋３声の声調変化や半３声、儿化には未対応である。声調変化や儿化については取扱説明書に記述があるものの、発音の仕方が実際のものと異なっていることは初級レベルの学習者によけいな混乱を招くことになりかねない。

✎ 6. 中国語以外の収録辞書

（次頁表参照）

✎ 7. まとめ

４機種のうち大学生向けに１機種推薦するとすれば、前号で田邉鉄氏も述べておられたようにやはり SR-T5030（SII）であろう。４機種の中では操作方法が比較的わかりやすく、英語辞書も充実している。しかしこれは総合的に見ればということであって、すでに述べてきたように SR-T5030 に問題点がないわけではけっしてない。まさしく一長一短であって電子辞書のどの機能に重きを置くのかによって選択は自ずと異なってくる。液晶は広くて見やすい PW-A8500（SHARP）が良いし、キーボードはキートップが大きくて打ちやすい SR-T5030（SII）が良い。中日辞典については２冊を搭載する PW-A8500（SHARP）。声調の入力は XD シリーズ（CASIO）方式が速い。音声付きが欲しければ XD-L シリーズ（CASIO）、簡体字手書き入力が必要なら Wordtank V70（Canon）だ。

昨今のメーカーの発売状況を見ると、中国語電子辞書はもはや珍しい変わり種ではなく、電子辞書のラインアップに定着したと思われる。機種を選べる時代にようやくなってきつつあるのだ。しかし英語と比べまだまだ収録される辞書のバリエーションが少なく、入力方法や検索方法にも改善の余地が大いにある。今後の充実に期待したい。

注

[1] 田邉鉄「学術リソースレビュー　電子辞書２機種」『漢字文献情報処理研究』第４号、2003 年

[2] http://www.casio.co.jp/exword/

[3] http://cweb.canon.jp/wordtank/index.html

[4] http://www.sharp.co.jp/e-dic/index.html

[5] http://www.sii.co.jp/cp/index.html

[6] １～３字まで入力可能。前方一致検索。

[7] XD-H7310 には用例・解説ボタンがあるが、Wordtank V70 は辞書本文を表示後、もう一度訳・決定ボタンを押すと例文が表示される。表示方法が直感的でなく、正直言って取扱説明書を見るまでわからなかった。

[8] 例えば“心”で終わる単語を検索するのであれば、「＊xīn」と入力する。

[9] XD-H7310 と PW-A8500 では「＊」ではなく「～」を使用。

[10] 小学館『中日辞典』の付録のうち、歴代王朝年表・中国語音節表・中国標準字形照合表を除いた部分を全て収録している。他の3機種は付録を収録していない。

[11] XD-H7310とWordtank V70は同じ会話集を収録するが、XD-H7310は基本表現と場面別会話のみ、Wordtank V70はそれに加えて辞書・資料編とコラム集を含む。

[12] 基本表現300には広東語・台湾語も含む。

[13] 中国語による検索はいずれも拼音による検索。

[14] 取扱説明書158ページ。XD-H7310に収録されている『ひとり歩きの中国語自由自在』（2001年）は未見だが、同書2004年版では確かに声調符号付きカタカナ表記のみで拼音はない。

[15] CASIOのニュースリリースによれば7300と7350は学生向け、7310と7360は会社員向けだそうである。7310・7360は、7300・7350に比べて英語系の辞書が減らされ、日本語系の辞書や百科事典（『マイペディア』）などが加えられている。
http://www.casio.co.jp/release/2004/xd_h7310_l7360.html
http://www.casio.co.jp/release/2004/exword.html

6. 中国語以外の収録辞書　一覧表

電子辞書	英語系	日本語系	その他
XD-H7310[15]	ジーニアス英和辞典 ジーニアス和英辞典 英語類語辞典 ロングマン現代アメリカ英語辞典 英文手紙用例事典 ビジネス英語自遊自在 英会話とっさのひとこと辞典	広辞苑・逆引き広辞苑 漢字源JIS版 言いえて妙なことば選び辞典 四字熟語辞典 故事ことわざ辞典 カタカナ語新辞典	マイペディア 家庭の医学 日経パソコン用語事典
Wordtank V70	ジーニアス英和辞典 ジーニアス和英辞典	無し	無し
PW-A8500	ジーニアス英和辞典 ジーニアス和英辞典 英語類語使い分け辞典 オックスフォード現代英英辞典 英文ビジネスレター事典 英会話Make it! 基本表現編 英会話Make it! 場面攻略編	広辞苑・逆引き広辞苑 漢字源JIS版	無し
SR-T5030	ジーニアス英和大辞典 ジーニアス和英辞典 オックスフォード現代英英辞典 コンサイスオックスフォード類語辞典 英会話とっさのひとこと辞典	広辞苑・逆引き広辞苑 漢字源JIS版 パーソナルカタカナ語辞典	無し

快譯通試用記

MD201 中英日空中学習辞典

小川　利康（おがわ　としやす）

✎快譯通とは

「快譯通」とは台湾の快譯通股份公司の製造する電子辞書のことで、台湾及び中国では相当なシェアを誇っており、デパートなどの電器製品売り場の片隅で必ず見かける製品である。普通は現地でないと購入できないものだが、最近はBIZASIA（http://www.asia-comic.com/shop/DTMD201-d.html）がネット上で通信販売をしており、日本国内でも手軽に購入することができた。

電子辞書としては、すでにカシオのEX-Word XD-R7300を持っていたので、中日・日中辞書が無くて困っているわけではなかったが、快譯通MD201にはオックスフォード英中辞典だけでなく、広東語・上海語会話入門（音声つき）など辞書以外のコンテンツも沢山盛り込まれているので、違った使い道がありそうだった。BIZASIAの売価は32,000円と決してお安くないが、パラレルポート経由でPCとリンクできるという機能やCFカードも使えるというあやしい魅力に抗しきれず、とうとう買ってしまった。以下、簡単に快譯通試用機をお届けする。

図1:「空中」とはFMを受信して学習できる為

国産電子辞書と比べる

今回の辞書特集で国内各社の電子辞書については清原氏によって詳論されているので、そちらを参照頂くとして、ここでは中日、日中辞書の語彙数について比較してみると、

●国内各社の電子辞書語彙数（何れも小学館版）

辞書名	収録語彙数
中日辞典第1版	85,000
日中辞典第1版	83,000

おおむねベースとしている紙版の辞書が同一なので、収録語彙数に大きな違いはない。これに対して、快譯通の場合は以下のようになる。

辞書名	収録語彙数
新編日華辞典	60,000
當代中日辞典	32,000

中日の方が弱いのは台湾では日本語から引かれるケースが多いことを想定しているためだろう。このデータだけを見たら、快譯通を購入する意味

は皆無に近い。小学館よりも新しい辞書であることは事実だが、新語が多いとも思えない。また、日本の辞書ほど例文がなく、決して学習者にとって語釈も親切ではない。通常の中国語学習者には国産の電子辞書を強く薦めたい。

多種多彩な辞書

快譯通の長所はむしろ多様な辞書を数多く収録しているところにある。代表的なところだけでも以下の辞書が収録されている。

英漢辞書（29 万語）
牛津（Oxford）当代大辞典（22 万語句）
中英辞書（21 世紀漢英大辞典：（6 万語）
外来語辞書（日本最新外来語辞典：3 万語句）
国語字典（収録数 1 万 3 千字）

MD201 はもともと台湾人向けであるから、この英語辞書機能がメインであることは当然のことでだが、注目に値するのは、単語音声の充実である。英中辞典に音声がついているのは驚くに値しないだろうが、漢英辞書から参照する中国語の単語についても全て北京語音・広東語音の音声が聞けるのである。人工合成音も含まれているが、自然音声も相当入っている。中日、日中辞典でも日本語、北京語ばかりか、広東語まで音が聞ける。香港人の北京語学習、台湾人の広東語学習を意識した設計で、我々日本人も恩恵にあずかれる得難い機能だ。

このほか台湾の辞書らしく、入力方法は多彩を極める。注音、五筆、倉頡、手書き入力など十二種類にのぼり、このほか日本語辞書に関してはローマ字、注音、平仮名などの入力方法が選べる。システム設定で文字表示も繁体字、簡体字を切り替えられる。繁体字をベースに開発したので、どちらでも文字表示は正確なようだ（面包，麵包などの区別）が、異体字の問題もあるので、繁体字で使った方が信頼性は高いだろう。

通常の電子辞書ならば、これで全て紹介が終わったことになるだろうが、これだけでは全内容の半分も紹介したことにならない。無料で添付される CD やネット上で公開されている電子書（1300 冊あるという）をインストールすると、大量のノウハウ本を読むことができる。その中でも多数を占めるのが語学入門書であることは言うまでもない。試みに幾つか紹介すると、

印傭會話便利通、家有泰傭立可通、家有越傭立可說、韓語入門、上海話一學就會、新新人類戀愛日語、新新人類流行日語、辦公室輕鬆日語、日語會話必備單字…

という具合に、これでもかという具合に詰め込まれ、おおむね音声付きである。日本でも書店に並んでいる語学入門書ほどのボリュームはなく、どれも半時あれば読み終わるようなものだが、このサービス精神には感服してしまう。このほかにも小説、占い、生活百科、グルメ、ゲームに至るまで、あらゆる分野がカバーされている。

さらにパラレルポート経由で接続し、MS Outlook 上の住所、メモ、予定を同期させることができる。当然データは全て Big5 でなければならないので、試みていないが、日本の 7,8 年前の PDA を彷彿とさせるインターフェイスで、使い勝手はそれなりである。ちょっとしたファイルならば、Outlook に放り込んでおけば、出先で参照することもできる。

国産のスマートな辞書と比べると液晶品質も悪く、日本人にとっては使い勝手も決して良いとは言えないので、万人にはお勧めできないが、極めて多機能な辞書であることは間違いない。チャンスがあれば是非触れてみて欲しい。

シンポジウム：「デジタル語学教材と著作権」

日本中国語 CAI 研究会主催公開シンポジウムの報告

山崎　直樹（やまざき　なおき）

✎ はじめに

日本中国語 CAI 研究会（会長：田邉　鉄）は、2003 年 11 月 30 日、中国語 CAI 研究会年次総会と併せて、標記のシンポジウムを開催した（於早稲田大学）。

これは、CAI 研究会の関係者が、この会に先立って（同年８月）開催された、漢字文献情報処理研究会主催のシンポジウム：『東洋学情報化と著作権問題』に参加し、このような研鑽の機会が非常に有意義なこと、さらに、外国語を教える教師に纏わる著作権問題に特化した研鑽の機会があれば、なお望ましいことを痛感し、実現させたものである。なお、講師は、漢情研主催のシンポジウムと同じく、慶應義塾大学産業研究所助教授・石岡克俊先生にお願いした。

✎ このシンポジウムの趣旨

以下に、このシンポジウムの案内文書に掲げた文章を転記し、このシンポジウムの趣旨説明としたい。

語学教師は、いろいろなものをコピーして、切り貼りして、プリントにして、教室で配付しているが、それは、どこまで許されるのか、「個人の視聴用に限って」というビデオを教室で上映していたりするが、それは、ほんとうに適法なのか、デジタルメディアを利用して同じことをしたばあいには、どう違ってくるのか……今後いっそう語学教材の電子化が進むと、デジタルゆえの複製の容易さが逆に新たな問題を生むことはないだろうか……

このシンポジウムではデジタル化に潜む著作権問題について、知的財産法の専門家である慶應義塾大学の石岡克俊先生をお迎えし、教材作りの現場からは山崎直樹先生、田邉鉄先生に具体的な問題提起をいただき、著作権の問題を考えたいと思います。

中国語に限らず、広く外国語教材の電子化に関心を寄せる方々の参加をお待ちしております。

✎ 事後雑感

このシンポジウムの開催にあたっては、CAI 研究会の関係者に対してのみではなく、各種教育機関で言語教育に携わっている人々にも、また、中国語関係の書籍を出版している出版社にも広報を行った。

その結果、当日の参加者は、言語教育を生業とする方々よりも、むしろ、出版関係者の方々のほうが多かった。参加していただいた出版関係者の方々には、まず、我々の会の趣旨をご理解いただいたことに感謝を申し上げたい。

また、同時に、出版方面においては、著作権の問題がいかに緊要なものであるかということを、主催者側が再認識する機会にもなった。

出版物という形態をとる語学教材の出来不出来に関しては、著者と出版社の関係は、良くも悪くも、共犯関係にあるのであるから（決して、主犯と従犯ではなかろう）、両者にとって、著作権の問題は、等しく緊要な問題であるはずなのだが。

✎『質問集』について

シンポジウムは､主催者側が用意した『質問集』に対し、逐次、講師から回答をいただく、という形式を採った。

以下に、その『質問集』のアウトラインを提示し､同様の催しをされる方々の参考に供したい(各地でこのような試みがなされることを切に希望する)。

なお、この『質問集』を最終的にまとめたのは、この文章の筆者（山崎）であるが、作成の過程で、CAI 研究会のメーリングリストにポストされた著作権関連の多くの質問を取り込ませていただいた。質問をお寄せくださった方々にも、感謝を申しあげたい。

また、『質問集』そのものをごらんになりたい方には、データファイル（フォーマット：PDF）をお送りするので、下記にご連絡いただきたい。ただし､『質問集』の中には､著作権的に見て､「危ない」画像なども多く含まれている（例：ウルトラマンの画像）ので、ご利用にあたっては、利用者自身で、その責を負われたい。

連絡先：

山崎直樹（大阪外国語大学）

ymzknk@osaka-gaidai.ac.jp

✎『質問集』の概要

Ⅰ　フォント

教室で使用するために、A4 サイズの紙に、単語をひとつずつプリントしたカードを作りました。同業者がそれを見て、われもわれもと欲しがるので、画像ファイルにして（どんな環境でも印刷できるようにとの配慮です）、WWW 上で公開することにしました。画像の元となった文字のフォントの著作権者との間に、問題は生じませんか？

Ⅱ　公共性の高いリソース

公共性の高い文書（例えば、新聞記事、法律の条文、入試問題……）は、教科書の一部に使うばあい、いちいち、許可をとっていないように思います。この慣習（？）は、WWW サイト上で使用すること前提に開発するデジタル教材の場合でも、通用するのでしょうか？

Ⅲ　自作の教材

WWW 上のリソースは、教育目的の名のもとに、どこまで、使用していいのでしょうか？

教師は、これまで、いろいろなもの切り貼りして、教材を作り、印刷して学生に配布してきました。

A:

ⅰ　例えば、日本相撲協会の WWW サイトからダウンロードした朝青龍や武蔵丸の画像を貼り付けて作った教材を、印刷して、学生に配布するのは、問題ないのでしょうか？

ⅱ　ⅰの画像を組み入れた電子形式の教材を作って、CD に焼いたのを、学生に配布するのはどうでしょう？

ⅲ　ⅱの教材を、学生の自習用に開放されている PC に置いて、誰もが使えるようにしておくのはどうでしょう？

B:

学生に対し、「MS-Word を使って、いろいろ

な『お知らせ』を作りなさい」という課題を出しました。学生は、WWW 上のあちこちのリソースから、いろいろな画像をダウンロードしてきて、それをぺたぺた貼り付けて、お知らせを作っています。もちろん、『お知らせ』の中では、出処を明示するなどということはできません。学生は、「学習成果発表会」で、嬉々として、その満艦飾のお知らせを展示しています。何か、釈然としないのですが……

IV　既存の教科書のデジタル化

質問（1）

紙媒体の教科書があります。これを丸ごとデジタル化する場合、最低限、誰の許可を取りつける必要がありますか？

質問（1-A）

質問⑴の個別の事例です。イラストなどは、出版社が買い取っていたとしても、イラストレーターなどから、「紙媒体での使用を前提として売り渡したのだ。デジタル化するなら、別途、許可を求めるべきだ」とクレームがつくおそれはありますか？

質問（1-B）

質問⑴の個別の事例です。語学教材には、CD などの形で音源が付きますが、その音源の吹き込み者から、同様のクレームがつく可能性は、ありますか？例：「CD として配付される形での使用を前提に売り渡したのだ……」

質問（1-C）

質問⑴の個別の事例です。ある教科書には、以下に掲げるような写真が載っています。この教科書を、そのままデジタル化するとしたら、どんなことに気をつける必要がありますか？

＜著名画家の作品＞＜映画の一場面＞＜紙幣＞

質問（2）

上記の⑴・(1-A)・(1-B)・⑵の問題は、デジタル化して使用するのが、スタンドアローン環境である場合と、ネットワーク環境（例えば、WWW サイト上）である場合とでは、何か異なる問題がありますか？

V　教科書の音源

わたしが作ったテキストを採用してくれた某先生から、次のような申し出が来ました。どうすべきでしょうか？

「テキストの課文を WWW サイトに掲載し、テキスト付属の音源から、一文ごとに音声を切り出したものを、リンクさせたい。音源が他に転用されるのが心配なら、ストリーム再生方式にして、ダウンロード＋保存ができないようにするが……」

VI　新しい形態の授業での利用

これまで「授業の過程で利用する」ことを目的に、文献やマルチメディア・リソースの複写や学生への配布を日常的に行ってきました。最近、遠隔授業やフリー・アクセスの授業など、新しい形態の授業が行われるようになっていますが、通常の授業と異なる配慮が必要でしょうか。例えば、次のようなケースではどうでしょうか。

⑴　ビデオで中国の映画を見せ、聞き取りさせる授業が、カリキュラム改編でフリー・アクセス化された。学生がいつでも授業を受けられるよう、中国映画の一部を VCD に焼いて配布してかまわないでしょうか。

⑵　⑴のケースで、VCD に焼くかわりに、学内のサーバに置き、LAN 上のパソコンで視聴できるようにしたらどうでしょうか。

⑶　他の大学とネットワークで協調授業を行う場合に、授業で輪読する書籍のコピーをサーバーに置いて共有することは問題ないでしょうか。